

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>5</sup> :</b> C12N 15/00, 15/12, 15/63 C12N 15/70, 15/74, 15/79	<b>A1</b>	<b>(11) International Publication Number:</b> WO 93/04166 <b>(43) International Publication Date:</b> 4 March 1993 (04.03.93)
<b>(21) International Application Number:</b> PCT/US92/06748 <b>(22) International Filing Date:</b> 17 August 1992 (17.08.92)  <b>(30) Priority data:</b> 746,032 16 August 1991 (16.08.91) US  <b>(71) Applicant:</b> CARNEGIE INSTITUTION OF WASHINGTON [US/US]; 1530 P Street N.W., Washington, DC 20005 (US).  <b>(72) Inventors:</b> MCKNIGHT, Steven, L. ; 7113 Oxford Road, Baltimore, MD 21212 (US). THOMPSON, Catherine, C. ; 1325 West 42nd Street, Baltimore, MD 21211 (US). LAMARCO, Karen, L. ; 1545 18th Street N.W. #708, Washington, DC 20036 (US).		<b>(74) Agents:</b> KOKULIS, Paul, N. et al.; Cushman, Darby & Cushman, 1615 L Street N.W., Washington, DC 20036 (US).  <b>(81) Designated States:</b> AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, SE).  <b>Published</b> <i>With international search report.</i>
<b>(54) Title:</b> PURINE-REGION DNA BINDING PROTEIN  <b>(57) Abstract</b>  The present invention relates to a GA binding protein and to DNA segments encoding subunits thereof.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	MN	Mongolia
AU	Australia	FR	France	MR	Mauritania
BB	Barbados	GA	Gabon	MW	Malawi
BE	Belgium	GB	United Kingdom	NL	Netherlands
BF	Burkina Faso	GN	Guinea	NO	Norway
BG	Bulgaria	GR	Greece	NZ	New Zealand
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IE	Ireland	PT	Portugal
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	RU	Russian Federation
CG	Congo	KP	Democratic People's Republic of Korea	SD	Sudan
CH	Switzerland	KR	Republic of Korea	SE	Sweden
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovak Republic
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CS	Czechoslovakia	LU	Luxembourg	SU	Soviet Union
CZ	Czech Republic	MC	Monaco	TD	Chad
DE	Germany	MG	Madagascar	TG	Togo
DK	Denmark	MI	Mali	UA	Ukraine
ES	Spain			US	United States of America

## PURINE-REGION DNA BINDING PROTEIN

5

BACKGROUND OF THE INVENTION

## Technical Field

The present invention relates, in general, to a DNA binding protein and to a DNA sequence encoding same. In particular, the invention relates to a GA binding protein and to DNA segments encoding the subunits thereof.

## Background Information

Herpes simplex virus 1 (HSV1) immediate early (IE) genes are induced at the outset of the lytic infection by a virion associated protein termed VP16 (Post et al, Cell 24, 555 (1981)). At least two classes of *cis*-regulatory elements qualify HSV IE genes for induction by VP16. The most essential VP16 *cis*-response element is characterized by the nonanucleotide sequence 5'-TAATGARAT-3' (Mackem et al, Proc Natl. Acad. Sci. U.S.A. 79, 4917 (1982); Mackem et al, J. Virol. 44, 939 (1982); Cordingley et al, Nucleic Acids Res. 11, 2347 (1983); Kristie et al, Proc. Natl. Acad. Sci. U.S.A. 81, 4065 (1984); Gaffney et al, Nucleic Acids Res. 13, 7847 (1985); Bzik et al, *ibid.* 14, 929 (1986); O'Hare et al, J. Virol. 61, 190 (1987); and Triezenberg et al, Genes Dev. 2, 730 (1988)). VP16 binds tightly to this DNA sequence in a complex with the cellular transcription factor Oct1 (Preston et al, C 11 52, 425 (1988); O'Hare et al, *ibid.* p. 435 (1988); and Gerster et al, Proc. Natl. Acad. Sci.

U.S.A. 85, 6347 (1988)). A second *cis*-regulatory element required for VP16-mediated induction of HSV IE genes consists of three imperfect repeats of the purine-rich hexanucleotide 5'-CGGAAR-3' (Triezenberg et al, Genes Dev. 2, 730 (1988) and Spector et al, *ibid.* 87, 5268 (1990)). A protein complex capable of avid interaction with the purine-rich repeats (GA repeats) has been identified in soluble preparations of rat liver nucleic (Triezenberg et al, Genes Dev. 2, 730 (1988)). This GA binding protein (GABP) consists of two separable subunits. Purified samples of either subunit do not interact with the GA repeats, yet regain potent DNA binding activity when mixed (LaMarco et al, Genes Dev. 3, 1372 (1989)).

Applicants have isolated cDNA clones encoding both subunits of GABP and have revealed that one (GABP $\alpha$ ) is related to the Ets transforming protein, while the other (GABP $\beta$ ) contains a series of 33-amino acid repeats related in sequence to a variety of proteins including Notch of Drosophila melanogaster, Lin12 and Glp1 of Caenorhabditis elegans and SW14 and SW16 of Saccharomyces cerevisiae (Wharton et al, Cell 43, 567 (1985); Greenwald, *ibid* 43, 583 (1985); Yochem et al, Nature 335, 547 (1988); Yochem et al, Cell 58, 553 (1989); Austin et al, *ibid.*, p. 565 (1989); Breeden et al, Nature 329, 651 (1987); and Andrews et al, *ibid.* 342, 830 (1989)). In addition, Applicants have demonstrated that these two protein sequence motifs, the Ets-related domain of GABP $\alpha$  and the 33-amino acid repeats of GABP $\beta$ , contribute the surfaces that form a multiprotein complex capable of stable and specific interaction with DNA. These findings, which form the basis of the present invention, provide insight into the problem of regulatory specificity and define, for the first time, a

discrete function for the 33-amino acid repeat motif.

#### SUMMARY OF THE INVENTION

5 It is a general object of the invention to provide DNA segments encoding the subunits of GABP.

In one embodiment, the present invention relates to a DNA segment encoding GABP $\alpha$ , GABP $\beta$ 1 or GABP $\beta$ 2, or portion thereof.

10 In a further embodiment, the present invention relates to a construct comprising at least one of the above-described segments and to a host cell transformed therewith.

Further objects and advantages of the present invention will become clear from the  
15 description that follows.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Amino acid sequences of tryptic peptides derived from GA binding proteins. GABP (20  $\mu$ g) was purified to homogeneity (inset) as described  
20 (LaMarco et al, Genes Dev. 3, 1372 (1989)) except that boiled salmon sperm DNA (20  $\mu$ g/ml) was included as a non-specific competitor in the DNA affinity chromatographic step. Approximately 500 picomoles of protein was lyophilized, reduced, acetylated, and  
25 subjected to cleavage by trypsin (Boehringer Mannheim). The resulting peptides were separated by reverse-phase HPLC as described (Stone et al, Laboratory Methodology in Biochemistry, Fini et al eds (RC Press, Boca Raton, FL, 1991)). Amino acid  
30 sequence analysis was performed on a Vydac C18 column (Appli d Biosystems 477-A). The amino acid sequences derived from peaks 1-13 were: 1, SLFDQGVIEK; 2, ?AWALEGY; 3, DEIS?VGDEGEFK; 4,

ELESINQEDFFQR; 5, LQESLDAHEIELQDIQL?P?R; 6,  
DQISIVGDEGEFK; 7, MAELV; 8, YVQASQLQQMNEIVTIDQP; 9,  
TPLHMWASEGHA; 10, GEILWS; 11, LIEIEIDGTEK; 12,  
ILMANGAPFTTD; 13, TGNNGQIQL?QFLLEL?TDR.

5     Figure 2. Nucleotide and deduced amino acid  
sequences of cDNAs encoding GABP subunits. (A)  
Sequence of GABP $\alpha$ . (B) Sequences of GABP $\beta$ 1 and  $\beta$ 2.  
An unamplified cDNA library prepared from mouse  
adipocyte mRNA was screened with a mixture of  
10     degenerate oligonucleotides derived from the amino  
acids sequences of peptides 3, 4, 5, and 8 (Fig. 1)  
labeled with  $^{32}$ P using polynucleotide kinase. The  
basic SSC protocol was used (Ausubel et al, Current  
Protocols in Molecular Biology (Wiley & Sons, NY),  
15     1989). Hybridization was performed at 48°C for  
16 hrs. GABP $\beta$ 1 and  $\beta$ 2 were isolated by screening a  
day-8.5 mouse embryo cDNA library (Lee, Mol.  
Endocrinol. 4, 1034 (1990)) with degenerate  
oligonucleotides corresponding to peptides 9 and 12  
20     (Fig. 1). Kinased oligonucleotide probes were  
hybridized in 6X SSC, 1X Denhardt's, 0.05% sodium  
pyrophosphate, and 100  $\mu$ g/ml yeast tRNA at 50°C for  
14 hours. Washing conditions were 6X SSC, 0.05%  
sodium pyrophosphate at 55°C. A total of five  
25     clones were isolated that hybridized with both  
oligonucleotide probes. Four of the clones were  
approximately 2.6 kb and differed only slightly in  
the length of the 5' untranslated region; these cDNA  
clones encoded GABP $\beta$ 1. The fifth cDNA clone was  
30     approximately 1.4 kb and differed from the other  
four at its 3' end; this cDNA clone encoded GABP $\beta$ 2.  
Four additional cDNA clones corresponding to GABP $\beta$ 2  
were subsequently identified. DNA sequencing was by  
the dideoxy chain termination method (Sanger et al,  
35     Proc. Natl. Acad. Sci. USA 74, 5463 (1977)) using  
Sequenase (U.S. Biochemicals) under conditions

suggested by the manufacturer. The complete nucleotide sequences of GABP $\alpha$ ,  $\beta$ 1 and  $\beta$ 2 were determined on both DNA strands using deleted templates or synthetic oligonucleotide primers.

5 Deletions were made using exonuclease III (Pharmacia) under conditions specified by the manufacturer. The sequences for GABP $\beta$ 1 and GABP $\beta$ 2 were identical up to nucleotide 1130 except for a three nucleotide insertion (GTA) at position 828 of

10 GABP $\beta$ 1. Sequencing of four other independent isolates of GABP $\beta$ 1 were identical to GABP $\beta$ 2 at this site. Peptides identified by amino acid sequencing of purified GABP are underlined in the deduced amino acid sequences. The dashed lines indicate the

15 sequence in GABP $\beta$ 1 not found in GABP $\beta$ 2. The sequence for  $\beta$ 2 is shown from the point at which it diverges from  $\beta$ 1.

Figure 3. Tissue distribution of GABP mRNAs. RNA was isolated from various rat tissues (Chingwin et al, Biochemistry 18, 5294 (1979)) and mouse L cells (Chomczynski et al, Anal. Biochem. 162, 156 (1987)).

20 10 $\mu$ g of poly A<sup>+</sup> RNA was separated on a 1% agarose-formaldehyde gel, transferred to Nytran (Schleicher and Schuell) and hybridized with a random-primed

25 probe prepared from GABP $\alpha$  (A) or GABP $\beta$ 1 (B).

Figure 4. Requirement of GABP $\alpha$  and GABP $\beta$ 1 for sequence specific DNA binding. A) In vitro translation of GABP proteins. Sense and anti-sense RNAs from GABP $\alpha$  and GABP $\beta$  were transcribed in vitro

30 and translated in rabbit reticulocyte lysates (Krieg et al, Nucl. Acids Res. 12, 7057 (1984)). Plasmids with cDNAs inserted in the Eco RI and Xho I sites of Bluescript (Stratagene) were linearized with Asp 718 and transcribed with T3 RNA polymerase to generate

35 sense strand RNA, or linearized with Bam HI and

transcribed with T7 RNA polymerase to generate anti-sense RNA. RNAs were used to program rabbit reticulocyte lysates in the presence of <sup>35</sup>S-methionine under conditions specified by the manufacturer (Promega Biotec). Unlabeled protein was used for DNA binding experiments. The <sup>35</sup>S-methionine labeled products were separated on a 12.5% SDS-polyacrylamide gel and visualized by fluorography; (-), anti-sense RNA; (+), sense strand RNA. Positions of molecular weight markers are indicated in kD. (B) Electrophoretic mobility shift assays with in vitro translated GABP proteins. Proteins were incubated in the presence of a <sup>32</sup>P-labeled DNA fragment from the HSV ICP4 promoter and subjected to electrophoresis on a non-denaturing 5% polyacrylamide gel in .5X TBE (Garner et al, Nucl. Acids Res. 9, 3047 (1981); Fried et al, *ibid*, p. 6505 (1981)). For DNA binding assays, samples containing in vitro translated protein were incubated in 25 mM Tris pH 8.0, 10% glycerol, 50 mM CK1, 3 mM MgCl<sub>2</sub>, 0.5mM EDTA, 1mM DTT, 50 µg/ml poly dIdC on ice for 10 minutes, then probe was added and incubation continued at room temperature for 10 minutes. The probe was a 180 bp Nco I-Sal I fragment excised from the herpes simplex virus ICP4 promoter. The fragment was labeled by fill-in with the Klenow fragment of DNA polymerase I in the presence of <sup>32</sup>P-dCTP. Protein:DNA complexes were subjected to electrophoresis on 5% (30:1) polyacrylamide gels in 0.5X TBE. Radioactive DNA and DNA:protein complexes were visualized by autoradiography. "B" indicates GABP bound DNA, "E" indicates DNA bound by proteins endogenous to reticulocyte lysates.

Figure 5. Schematic diagram of GABP subunits showing regions of amino acid sequence similarity to



related proteins. (Top) GABP $\alpha$  is represented as a rectangle with the NH<sub>2</sub>-terminus on the left and the COOH-terminus on the right. The region of sequence similarity to Ets-related proteins is shaded (amino acids 316-400) and compared with the sequences of Ets-1 (Gunther et al, Genes Dev. 4, 667 (1990)), Erg (Reddy et al, Proc. Natl. Acad. Sci. U.S.A. 84, 6131 (1987)), Elk (Rao et al, Science 244, 66 (1989) and E74A (Burtis et al, Cell 61, 85 (1990)). Residues that are common to GABP $\alpha$  and other proteins are boxed in black. (Bottom) GABP $\beta$ 1 is represented as a rectangle with the NH<sub>2</sub>-terminus on the left and the COOH-terminus on the right. The 33-amino acid repeats are shown as shaded rectangles. The unique COOH-terminal segment of GABP $\beta$ 1 relative to GABP $\beta$ 2 is indicated in black (333-382). The sequence of the four 33 amino acid repeats in GABP $\beta$ 1 are shown below; residues that are common to two or more repeats are boxed in black and used to derive the GABP $\beta$  consensus. Similar criteria were used to derive consensus sequences for the 33 amino acid repeats of cdc 10/SW14,6 (Ares et al, EMBO J. 4, 457 (1985); Andrews et al, Nature, 342, 830 (1989); Breeden et al, Nature 329, 651 (1987)), Notch (Wharton et al, Cell 43, 567 (1985); Greenwald, *ibid.* 583 (1985)), glp1 (Yochem et al, Cell 58, 553 (1989); Austin et al *ibid.* p. 565 (1989)), lin12 (Yochem et al, Nature 335, 547 (1988)), ankyrin (Lux et al *ibid.* 344, 36 (1990)), NF B Kieran et al, Cell 62, 1007 (1990); Bours et al, Nature 348, 76 (1990)), fem1 (Yochem et al, Cell 58, 553 (1989); Austin et al *ibid.*, p. 565 (1989)), and bcl-3 (Ohno et al *ibid.*, p. 991 (1990)). The repeats from cdc10, SW14 and SW16 were combined to determine the consensus. The consensus for ankyrin was taken from Lux et al, Nature 344, 36 (1990). The overall consensus was defined as residues present in at

least 6 of the individual consensus repeat sequences.

Figure 6. DNA binding by GABP expressed in bacteria. Purified proteins were incubated with a <sup>32</sup>P-labeled oligonucleotide containing the GABP binding site derived from the enhancer of the herpes simplex virus ICP4 gene (5'-TGCGGAACGGAAGCGGAAACCGCCGGATCG-3') (Triezenberg et al, Genes Dev. 2 (1988); LaMarco et al, *ibid.* 3, 1372 (1989)). Free and protein-bound DNA samples were subjected to electrophoresis on 5% polyacrylamide gels in either 0.5X TBE (A) or 0.25X TBE buffer (B).

Figure 7. Characterization of the DNA binding site for GABP. (A) Increasing concentrations of GABP $\alpha$ , either in the absence (left panel) or presence (right panel) of GABP $\beta$ 1 were mixed with a <sup>32</sup>P-labeled DNA fragment derived from the herpes simplex virus ICP4 enhancer. Free and protein-bound complexes were partially digested with DNase I and subjected to electrophoresis on an 8% polyacrylamide sequencing gel. The positions of three purine-rich repeats within the region of DNA protected from digestion by GABP are indicated by arrows. Lanes 1-6 (left panel) show digestion patterns resulting from GABP $\alpha$  concentrations starting at 1.5 nM and decreasing in 3-fold increments to 0.005 nM. Lanes 1-6 (right panel) show patterns resulting from addition of the same concentrations of GABP $\alpha$  that had been supplemented with 0.5 nM of GABP $\beta$ 1. (B) Methylation protection (left panel) and interference (right panel) assays of DNA binding by GABP. The same DNA fragment used in (A) was incubated with GABP $\alpha$ , GABP $\beta$ 1, or an equimolar mixture of the two subunits, and exposed to dimethyl sulfate (DMS).

Partially methylated DNA was recovered, cleaved with piperidin , and run on an 8% polyacrylamide sequencing gel. For methylation interference assays, DNA was partially methylated, incubated with  
5 an equimolar mixture of GABP $\alpha$  and GABP $\beta$ 1, and subjected to electrophoresis on a non-denaturing polyacrylamide gel as described in Fig. 6. Free and protein bound DNA species were recovered, cleaved with piperidine, and electrophoresed on an 8%  
10 polyacrylamide sequencing gel. Nucleotide residues closely contacted by GABP are shown in the lower part of (B). Filled circles identify guanine residues that were protected from DMS by GABP. Methylation of the same four guanine residues also  
15 inhibited DNA binding by GABP. Open circles identify adenine residues where methylation is enhanced in the presence of both GABP $\alpha$  and GABP $\beta$ 1.

Figure 8. Measurements of DNA binding stability of complexes formed by various mixtures of GABP  
20 subunits. A  $^{32}$ P-labeled oligonucleotide containing a GABP binding site (Fig. 6) was incubated with GABP $\alpha$  alone, or together with equimolar amounts of either of the two  $\beta$  subunits. After a 10 minute incubation at 24°C, protein:DNA complexes were challenged with  
25 a 500-fold excess of unlabeled oligonucleotide. Protein-bound and free DNA were separated by non-denaturing gel electrophoresis as described in Fig. 6A. Protein-bound and free DNAs were located by autoradiography, excised, and quantitated by  
30 scintillation spectrometry. Results are presented as fraction of probe bound, normalized to 1.0 at the start point (time = 0).

Figure 9. UV-mediated crosslinking of GABP subunits to DNA. Isolated or mixed GABP subunits were  
35 incubated with a  $^{32}$ P-labeled oligonucleotide

containing a GABP binding site (Chodosh in Current  
Protocols in Molecular Biology, Vol. II, Ausubel et  
al eds (Greene Wiley, New York, 1988)) then exposed  
to ultraviolet light for varying lengths of time.

5 UV crosslinking was performed using an  
oligonucleotide composed of a GA binding site  
flanked by 10 bp of non-specific sequence (5'  
AACCAAGCTTGC GGAACGGAAGCGGAAACCG 3') corresponding to  
residues located between 280 and 300 bp upstream of  
10 the herpes simplex virus gene encoding ICP4.  
Oligonucleotides were labeled to high specific  
activity by fill-in reaction with the Klenow  
fragment of DNA polymerase I in the presence of all  
four <sup>32</sup>P-labeled dNTPs. DNA binding reactions were  
15 performed as described in a 96-well culture dish,  
followed by exposure to ultraviolet light. Samples  
were boiled in SDS-sample buffer and subjected to  
electrophoresis on SDS-polyacrylamide gels.  
Crosslinked protein species were visualized by  
20 autoradiography. Samples were denatured by boiling  
in SDS sample buffer and subjected to  
electrophoresis on a denaturing 12.5% polyacrylamide  
gel. Following electrophoresis the gel was dried  
and exposed to X-ray film. Time of exposure to UV  
25 light (minutes) is indicated above each gel lane.  
Migration of molecular weight marks (kD) is shown on  
the left.

Figure 10. Glutaraldehyde crosslinking of GABPβ1  
and GABPβ2 subunits. Bacterially synthesized  
30 proteins were incubated in phosphate buffered saline  
with varying concentrations of glutaraldehyde as  
indicated below each lane for five minutes at room  
temperature. Samples were denatured by boiling in  
SDS sample buffer and subject to electrophoresis on  
35 a denaturing 10% polyacrylamide gel. Following  
electrophoresis the gel was stained with Coomassie

brilliant blue. Proteins present in crosslinking reactions are indicated above each lane. BN110 is a truncated version of GABPB1 missing 110 NH<sub>2</sub>-terminal residues (see Fig. 12B).

5 Figure 11. DNA binding and complex formation assays of deleted variants of GABP $\alpha$ . Top panel shows  
schematic representation of GABP $\alpha$  deletion mutants. Individual mutants are designated according to the  
10 position of deletion end points with respect to the amino acid sequence of GABP $\alpha$ . Prefix "N" designates  
deletions missing residues starting at the NH<sub>2</sub>-terminus of GABP $\alpha$ , prefix "C" designates deletions  
missing COOH-terminal residues, numbers indicate the position of the amino acid at which the deletion  
15 terminates. The Ets-related segment of GABP $\alpha$  is highlighted by grey stippling. Bottom panel shows  
an autoradiographic image of a non-denaturing gel used to separate DNA:protein complexes formed  
between variants of GABP $\alpha$ , GABPB1 and a <sup>32</sup>P-labeled oligonucleotide that contained a GABP binding site.  
20 Each variant of GABP $\alpha$  was tested for DNA binding in the absence and presence of GABPB1 as indicated  
above the individual lanes.

Figure 12. Complex formation and UV crosslinking  
25 assays of deleted variants of GABPB1. Top panels of (A) and (B) show schematic representations of GABPB1  
deletion mutants. Individual mutants are designated according to the positions of deletion end points  
with respect to the amino acid sequence of GABPB1. Prefix "N" designates deletions from the NH<sub>2</sub>-  
30 terminus of GABPB1 (B), prefix "C" designates deletions missing COOH-terminal residues. Repeated  
sequences 33 or 32 amino acids in length that are related to similarly sized repeats in the Notch  
35 protein of Drosophila melanogaster are highlighted

by grey stippling. Unique parts of GABPB1 and GABPB2 are indicated by black and hatched rectangles at their respective COOH-termini. Deleted formation with GABP $\alpha$  as shown in the lower left panels of (A) and (B). Each deletion mutant was also tested in UV crosslinking assays shown in the lower right panels of (A) and (B). All complex formation and UV crosslinking assays were conducted in the presence of GABP $\alpha$  and a <sup>32</sup>P-labeled oligonucleotide containing a GABP binding site.

Figure 13. Model depicting complex formed between GABP and DNA. The sequence of the GABP binding site consists of two hexanucleotide repeats of the sequence 5'-CGGAAR-3' as in lower part of Fig. 13. Oval spheres directly above guanine residues of each hexanucleotide correspond to GABP $\alpha$  subunits, elongated rectangles correspond to 33 amino acid repeats of GABPB subunits. Smaller rectangles shown at top correspond to the region of GABPB1 required for formation of stable homodimers. Circular arrows designate flexible regions inferred to occur between the dimer forming region of GABPB1 and the 33-amino acid repeats located at its NH<sub>2</sub>-terminus.

#### DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to a DNA segment encoding all (or a unique portion) of the heteromeric transcriptional regulatory protein termed GA binding protein (GABP). The invention further relates to the encoded proteins (or polypeptides). A "unique portion" as used herein consists of at least five (or six) amino acids or, correspondingly, at least 15 (or 18) nucleotides. The present invention further relates to a

recombinant DNA molecule comprising the above DNA segment and to host cells transformed therewith.

In particular, the present invention relates to a DNA segment that encodes the entire amino acid sequence of GABP $\alpha$ , GABP $\beta$ 1 or GABP $\beta$ 2 given in Figure 2 (the specific DNA segments given in Figure 2 being only examples), or any unique portion thereof. DNA segments to which the invention relates also include those encoding substantially the same protein subunits as shown in Figure 2, including, for example, allelic and species variations thereof and functional equivalents of the amino acid sequences of Figure 2. The invention further relates to a DNA segment substantially identical to one of the subunit sequences shown in Figure 2. A "substantially identical" sequence is one the complement of which hybridizes to one of the sequences of Figure 2 at 50°C and 6X SSC (saline/sodium citrate) and which remains bound when subjected to washing at 55°C with 6X SSC (note: 20 x SSC = 3M sodium chloride/0.3 M sodium citrate). The invention also relates to nucleotide fragments complementary to such DNA segments. Unique portions of the DNA segment, or complementary fragments, can be used as probes for detecting the presence of respective complementary strands in DNA (or RNA) containing samples.

The present invention further relates to GABP, and subunits thereof, substantially free of proteins with which it is normally associated, and more especially, to unique peptide fragments of the subunits of that protein. The GABP protein (or functionally equivalent variations thereof), or peptide fragments thereof, to which the invention relates, also includes those which are chemically synthesized using known methods. The proteins and

peptides of the present invention can be modified, for example, phosphorylated, or unmodified.

The present invention also relates to recombinantly produced GABP, or subunits thereof, 5 having the amino acid sequence shown in Figure 2 or functionally equivalent variation thereof. The recombinantly produced protein can be modified, for example phosphorylated, or unmodified. The present invention, more particularly, relates to 10 recombinantly produced unique peptide fragments of GABP subunits.

The present invention also relates to a recombinant DNA molecule (or construct) and to a host cell transformed therewith. Using standard 15 methodologies, well known in the art, a recombinant DNA molecule comprising a vector and a DNA segment encoding at least one GABP subunit, or a unique portion thereof, can be constructed. Vectors suitable for use in the present invention include 20 plasmid and viral vectors. The vector can be selected so as to be suitable for transforming prokaryotic or eukaryotic cells. Advantageously, the recombinant molecule includes a promoter operably linked to the GABP encoding segment.

25 The recombinant DNA molecule of the invention can be introduced into appropriate host cells by one skilled in the art using method well known in the art. Suitable host cells include prokaryotic cells, such as bacteria, lower 30 eukaryotic cells, such as yeast, and higher eukaryotic cells, such as mammalian cells. These cells can serve as a source of GABP when cultured under appropriate conditions.

As noted at the outset, and as will be 35 further described in the Examples that follow, significant amino acid sequence similarity exists between GABP $\alpha$  and the products of the *est1* and *ets2*



(Watson et al, Proc. Natl. Acad. Sci. U.S.A. 85, 7862 (1988); Gunther et al, Genes Dev. 4, 667 (1990)) proto-oncogenes. The Ets-related region of GABP $\alpha$  is located close to the COOH-terminus of the subunit. Biochemical studies of Ets1 (Gunther et al, Genes Dev. 4, 667 (1990)), as well as the related proteins PU (Klemsz et al, Cell 61, 113 (1990)) and E74 (Urness et al, Cell 63, 47 (1990)), have demonstrated direct, sequence-specific DNA binding. These proteins, as well the products of several additional eukaryotic genes, share sequence similarity in an 85-amino acid region that is required for DNA binding (Karim et al, Genes Dev. 4, 1451 (1990)). The region of GABP $\alpha$  that is related to this family of proteins coincides with the 85 amino acid DNA binding domain (Fig. 5).

The amino acid sequences of GABPB1 and GABPB2 contain four repeats of a related amino acid sequence located at the NH<sub>2</sub>-termini of both subunits (Fig. 5). The first two repeats are 32 amino acids in length and the second two contain 33 amino acids. Similar repeats occur in the Notch protein of Drosophila melanogaster (Wharton et al, Cell 43, 567 (1985); I. Greenwald, *ibid.*, 583 (1985)), and the Lin12 and Glp1 proteins of Caenorhabditis elegans (Yochem et al, Nature 335, 547 (1988), Yochem et al, Cell 58, 553 (1989); Austin et al, *ibid.*, p. 565 (1989)). These "33-amino acid repeats" were first recognized in studies of the yeast protein SW16, which regulates gene expression involved in mating type switching (Breeden et al, Nature 329, 651 (1987)). Similar repeats have been identified in ankrin, a multifunctional protein associated with the membrane of red blood cells (Lux et al, Nature 344, 36 (1990)), several vaccinia virus encoded proteins of unknown function (Gillard et al, Proc. Natl. Acad. Sci. U.S.A. 83, 5573 (1986)), and the

transcription factor NF $\kappa$ B (Kieran et al, Cell 62, 1007 (1990); Bours et al, Nature 348, 76 (1990)).

5 The two subunits of GABP exhibit primary sequence motifs typical of proteins normally found in different cellular compartments. Accordingly, transcriptional regulatory proteins, such as members of the Ets family, might interact with membrane bound proteins that contain the 33-amino acid repeats present in GABP $\beta$ . The Notch, Glp1 and Lin12  
10 proteins might sequester transcription factors at the plasma membrane which could be released in response to appropriate extracellular signaling events. Alternatively, the cytoplasmic segments of these transmembrane proteins might be proteolyzed in  
15 response to an extracellular signal, allowing the 33-amino acid repeats to be translocated to the nucleus where they could abet the action of a second subunit. Either scenario would offer a direct pathway of signal transduction.

20 The results set forth in the Examples that follow demonstrate the reliance of competent DNA binding complexes on multiple subunits. By separating functional components onto different polypeptide chains, critical subunits might be  
25 differentially expressed or sequestered, generating useful strategies for regulation. For example differential expression of the mRNAs encoding the  $\beta$ 1 and  $\beta$ 2 subunits of GABP would be expected to impact substantially on the function of the resulting  
30 complex. In this regard, it is interesting to note that in cells undergoing replication, the  $\beta$ 2 subunit predominates whereas in non-dividing cells, subunit  $\beta$ 1 predominates.

35 It will be clear to one skilled in the art from a reading of this disclosure that advantage can be taken of information provided herein to effect alterations of both normal and abnormal expression

patterns regulated by the binding complexes described above. Such alterations can be effected, for example, using a variety of gene therapy protocols. It is contemplated that by altering the relative amounts of the  $\beta 1$  and  $\beta 2$  subunits, disease states characterized by rapid cell division, for example, cancer, can be controlled.

Certain aspects of the invention are described in greater detail in the non-limiting Examples that follow.

### Example 1

#### Isolation of Recombinant cDNA Clones

GABP (20  $\mu$ g) was purified from rat liver nuclear extracts and cleaved with trypsin. Proteolyzed fragments were separated by high performance liquid chromatography (HPLC), recovered, and subjected to gas-phase amino acid sequencing (Fig. 1). Partial sequences were derived from 13 tryptic peptides. Degenerate oligonucleotides capable of encoding four of the thirteen peptide sequences were synthesized and used as hybridization probes to screen an adipocyte cDNA library. Degenerate oligonucleotides were labeled with  $^{32}$ P using polynucleotide kinase. The basic sodium chloride/sodium citrate (SSC) protocol was used for screening (F.M. Ausubel et al, Current Protocols in Molecular Biology (Wiley and Sons, NY), 1989). Hybridization was performed at 48°C for 16 hours. One recombinant bacteriophage that contained a cDNA insert of 2 kb gave a positive signal when hybridized with each of the four oligonucleotide probes.

The insert of this recombinant was sequenced and found to contain an opening reading

frame that encoded a protein of 454 amino acids (Fig. 2A). The predicted molecular weight of this polypeptide (51.3 kD) corresponded to the size of the GABP $\alpha$  subunit purified from rat liver nuclei (LaMarco et al, Genes Dev. 3, 1372 (1989); Fig. 1). Inspection of the deduced amino acid sequence revealed segments that corresponded to eight of the 13 peptides isolated by trypsin digestion of intact GABP. On the basis of the latter two observations, this 454 residue polypeptide was tentatively identified as GABP $\alpha$ .

Degenerate oligonucleotides capable of encoding two of the tryptic peptide sequences not present in GABP $\alpha$  were synthesized and used as hybridization probes to search for a cDNA clone that encoded GABP $\beta$  (Lee, Mol. Endocrinol. 4, 1034 (1990)). Kinased oligonucleotide probes were hybridized in 6X SSC, 1X Denhardt's, 0.05% sodium pyrophosphate, and 100  $\mu$ g/ml yeast tRNA at 50°C for 14 hours. Washing conditions were 6X SSC, 0.05% sodium pyrophosphate at 55°C. A total of five clones were isolated that hybridized with both oligonucleotide probes. Four of the clones were approximately 2.6 kb and differed only slightly in the length of the 5' untranslated region; these cDNA clones encoded GABP $\beta$ 1. The fifth cDNA clone was approximately 1.4 kb and differed from the other four at its 3' end; this cDNA clone encoded GABP $\beta$ 2. Four additional cDNA clones corresponding to GABP $\beta$ 2 were subsequently identified. Five recombinant bacteriophage were identified according to their capacity to hybridize with both oligonucleotide probes. One of the cDNA clones differed at the 3' end from the other four. The largest cDNA insert of the four (2.6 kb) and the variant (1.4 kb) were sequenced (Fig. 2B). Both DNA sequences revealed long open reading frames specifying highly similar

polypeptid s. One cDNA encoded a protein of 382 amino acids, the other encoded a protein of 349 residues. Starting at their respective NH<sub>2</sub>-termini, the two proteins exhibited identical sequences for 333 amino acids. At this point the sequences to the two protein diverged such that the longer one contained an additional 50 residues before its terminus. The divergent COOH-terminal segments bore no apparent amino acid sequence similarity. The open reading frames of both polypeptides contained segments that corresponded to the two tryptic peptides used to design hybridization probes. Moreover, the predicted molecular weights of the two polypeptides (41.3 and 37 kD) corresponded closely with the size of the GABP $\beta$  subunit purified from rat liver nuclei (LaMarco et al, Genes Dev. 3, 1372 (1989); Fig. 1). The 41 kD polypeptide was therefore provisionally designated as GABP $\beta$ 1 and the 37 kD polypeptide as GABP $\beta$ 2.

20

### Example 2

#### Tissue Distribution of mRNA Encoding GABP $\alpha$ , GABP $\beta$ 1 and GABP $\beta$ 2

Northern (RNA) blot assays were used to determine the sizes and tissue distributions of mRNA encoding GABP $\alpha$ , GABP $\beta$ 1 and GABP $\beta$ 2 (Fig. 3). The cDNA corresponding to GABP $\alpha$  identified three mRNAs of roughly 5.0, 2.8 and 2.6 kb, which were expressed in a variety of tissues. The GABP $\alpha$  cDNA, which consisted of slightly less than 2.0 kb (Fig. 2A), represents a partial copy of any of the three mRNAs. Two mRNAs measuring 2.7 and 1.5 kb were identified in northern blots probed with GABP $\beta$ 1 cDNA. Like GABP $\alpha$  mRNAs, those encoding GABP $\beta$ 1 had a wide tissue distribution. Because the cDNAs that encoded GABP $\beta$ 1

and GABPB2 measured 2.6 and 1.4 kb, respectively (Fig. 2B), they probably represent nearly full-length copies of the respective mRNAs. Moreover, because the nucleotide sequences of the two cDNAs are identical from their respective 5' termini to the point of abrupt divergence 1.1 kb internal to the mRNA, they likely represent alternatively spliced transcripts derived from the same gene. Consistent with this interpretation is the presence of a potential splice donor site (AG dinucleotide) immediately preceding the point of divergence.

### Example 3

#### GABP DNA Binding Activity

To test whether the recombinant DNA clones described above possessed GABP DNA binding activity, reticulocyte lysates were programmed with RNA synthesized from the cDNAs that encode GABPB, GABPB1 and GABPB2. Each RNA was translated to form a protein product of the expected size (Fig. 4A). Individual lysates or mixtures thereof were tested for DNA binding to a fragment from the HSV1 ICP4 promoter that contained three GA repeats. Protein:DNA mixes were subjected to electrophoresis on nondenaturing polyacrylamide gels to separate free DNA from that complexed with protein. Reticulocyte lysate that had not been programmed with exogenous RNA contained protein(s) capable of forming a complex with the oligonucleotide probe that migrated more rapidly than the complex formed by GABP. Other than background activity endogenous to the reticulocyte lysate, specific protein:DNA complexes were not observed when lysates programmed with GABP $\alpha$ , GABPB1, or GABPB2 were tested in electrophoretic mobility shift assays. Likewise, no

new DNA binding activity was observed with lysate that had been used to co-translate RNAs encoding GABP $\alpha$  and GABP $\beta$ 2. However, co-translation of RNAs encoding GABP $\alpha$  and GABP $\beta$ 1 caused the lysate to form a DNA binding activity that could be distinguished from background (Fig. 4B). The interdependency of GABP $\alpha$  and GABP $\beta$ 1 observed in these assays is consistent with earlier observations that tested subunits purified from rat liver nuclei (LaMarco et al, Genes Dev. 3, 1372 (1989)).

#### Example 4

##### DNA Binding Properties of GABP

Recombinant cDNA copies of the mRNAs that encode GABP $\alpha$  and GABP $\beta$ 1 were introduced into bacteriophage T7 based vectors that allowed synthesis of the corresponding proteins in Escherichia coli (Studier et al, J. Mol. Biol., 189, 113 (1986)). Polymerase chain reaction was used to introduce a Bam HI site at the 5' end of the open reading frames encoding GABP $\alpha$  or GABP $\beta$ 1. cDNAs lacking the 3' untranslated region were inserted into a modified pT5 vector, which adds two amino acids (gly-ser) at the NH<sub>2</sub>-terminus of the encoded protein. Each subunit was expressed and purified using conventional chromatographic techniques. Proteins were expressed in bacteria as described (Shuman et al, Science 249, 771 (1990)). GABP $\alpha$  was precipitated from the soluble fraction by the addition of one volume of 2M ammonium sulfate in buffer A (10 mM Tris-HCl, pH 7.9, 100 mM KCl, 0.5 mM EDTA, 1 mM MgCl<sub>2</sub>, 2 mM DTT, 0.1 mM phenylmethylsulfonyl fluoride) with 2 mM CaCl<sub>2</sub>. The ammonium sulfate pellet was resuspended in 25 ml buffer B (25 mM Tris-HCl, pH 8.0, 0.75 mM EDTA, 10%

(v/v) glycerol, 1 mM DTT) with 75 mM NaCl and dialyzed against the same buffer. The dialysate was loaded onto a column of Q-Sepharose Fast Flow (Pharmacia). GABP $\alpha$  was eluted with a 75-500 mM NaCl gradient in buffer B. Peak fractions were pooled, dialyzed against buffer B and loaded onto a salmon sperm DNA-sepharose column. GABP $\alpha$  was eluted with a 0-400 mM NaCl gradient. GABP $\alpha$  was judged by Coomassie Blue staining of SDS polyacrylamide gels to account for greater than 90% of the total protein. GABP $\beta$ 1 was solubilized from the particulate fraction of bacterial extracts by sonication in buffer A supplemented with 7 M urea. The urea solubilized fraction was dialyzed against buffer B with 75 mM NaCl and centrifuged at 16,300 x g for one hour. The supernatant was applied to a Q-Sepharose column and eluted with a gradient of 75-500 mM NaCl. GABP $\beta$ 1 was judged to account for greater than 90% of total protein by Coomassie Blue staining of SDS-polyacrylamide gels.

The DNA binding properties of the two individual polypeptides and mixtures thereof were first studied by gel retardation using a DNA substrate derived from the enhancer of an immediate early gene of herpes simplex virus. Consistent with earlier studies (LaMarco et al, Genes Dev., 3, 1372 (1989)), binding was not observed when DNA was incubated with either of the isolated subunits. When GABP $\alpha$  and GABP $\beta$ 1 were incubated with DNA simultaneously, a DNA:protein complex exhibiting substantially retarded mobility relative to free DNA was observed (Fig. 6, left panel).

The multi-subunit dependence of DNA binding by GABP was relieved in gel retardation assays conducted at lower ionic strength (Fig. 6, right panel). GABP $\alpha$  formed two retarded complexes that migrated at positions between free DNA and the



complex formed with both subunits. Under these conditions, the mixture of GABP $\alpha$  and GABP $\beta$ 1 again led to the formation of a slowly migrating DNA:protein complex. The retarded mobility of the latter complex, relative to those formed by GABP $\alpha$  alone, reflected the presence of the GABP $\beta$ 1 subunit. The validity of this interpretation was confirmed by the use of antisera specific to each subunit. Antiserum specific to GABP $\alpha$  further retarded the migration of complexes formed with GABP $\alpha$  alone or the mixture of GABP $\alpha$  and GABP $\beta$ 1. Antiserum specific to GABP $\beta$ 1 did not affect the mobility of complexes formed between GABP $\alpha$  DNA, but retarded the complex formed in the presence of both subunits. Polyclonal antisera were generated by injecting rabbits with purified GABP $\alpha$  or GABP $\beta$ 1. Antisera were added to gel shift reactions at a dilution of 1:20. Pre-immune sera did not effect the migration of protein:DNA complexes.

The HSV1-derived DNA fragment used in binding assays of GABP contains three imperfect repeats of the hexanucleotide sequence 5'-CGGAAR-3', which were shown in earlier studies to be protected from DNase I digestion when bound by GABP (Triezenberg et al, Genes Dev. 2 (1988); LaMarco et al, *ibid.* 3 1372 (1989)). DNase I footprinting assays were performed using bacterially synthesized proteins under conditions that allowed interaction of GABP $\alpha$  alone. As shown in Fig. 7A, GABP $\alpha$  was capable of protecting the repeated hexanucleotide motifs from DNase I digestion when added at a concentration of 0.15 nM. When the GABP $\beta$ 1 subunit was added, protection was observed at a 10-fold more dilute concentration of GABP $\alpha$  (0.015 nM). In addition, the pattern of nuclease protection was extended slightly beyond the adenine residues of the third repeat.

The segment of DNA protected from DNase I digestion by GABP encompassed all three hexanucleotide repeats, yet was not centered over the repeats. Methylation protection and interference assays were undertaken in order to gain a more refined image of the sites of close contact established between GABP and DNA. Methylation protection assays conducted with GABP $\alpha$  showed a pattern of protection that included both guanine residues of the second and third hexanucleotide repeats. The same sets of guanines were protected when GABP $\beta$ 1 was added to the binding reaction. In the latter case, however, accentuated methylation was observed at adenine residues located adjacent to the guanine dinucleotides of the second and third repeats. Sites of methylation interference were mapped by separating protein-bound DNA molecules from those inactivated by partial methylation. Methylation of guanine dinucleotides in the second and third hexanucleotide repeats inhibited binding by the mixture of GABP $\alpha$  and GABP $\beta$ 1 (Fig. 7B).

The results of methylation protection and interference assays indicate that GABP binds to sites on DNA corresponding to two of the three purine-rich hexanucleotide repeats. The pattern of protection of guanine residues by GABP $\alpha$  was similar to that observed with the mixture of both subunits, indicating that GABP $\alpha$ , when added at a sufficiently high concentration, can bind specifically to DNA in the absence of GABP $\beta$ 1. Such observations offer an explanation for the two retarded bands observed when DNA was challenged with GABP $\alpha$  alone under conditions of low ionic strength (Fig. 6, right panel). The less retarded of the two bands is interpreted to represent a complex wherein GABP $\alpha$  is associated with only one of the two hexanucleotide repeats, while the more retarded complex is interpreted to contain

GABP $\alpha$  subunits associated with two hexanucleotide repeats. Binding assays that test d DNA probes containing a single hexanucleotide repeat supported this interpretation. When incubated with GABP $\alpha$  and assayed in low ionic strength gels, such DNA probes generated only one retarded complex.

Several observations indicate that the mixture of GABP $\alpha$  and GABP $\beta$ 1 forms a complex that binds DNA more stably than the  $\alpha$  subunit alone (LaMarco et al, Genes Dev. 3, 1372 (1989)). To further investigate the effect of GABP $\beta$ 1 on DNA binding by GABP $\alpha$ , the rate at which variously mixed proteins dissociate from DNA was measured (Fig. 8). The dissociation rate of GABP $\alpha$  alone was too rapid to be accurately measured. Less than ten percent of the DNA remained bound to GABP $\alpha$  after a 10 second challenge with excess, unlabeled competitor DNA. In contrast, when both GABP $\alpha$  and GABP $\beta$ 1 were present, the dissociation rate was much slower ( $T_{1/2}$  = 1.5 min). Similar assays were performed with a mixture of GABP $\alpha$  and GABP $\beta$ 2, which, in earlier experiments, failed to form a stable complex with DNA. When used at nM concentrations, GABP $\beta$ 2 was capable of forming a DNA binding complex with GABP $\alpha$  (see Fig. 12A). The  $\beta$ 2 isoform of GABP also stabilized DNA binding by GABP $\alpha$ , yet yielded a complex that dissociated more rapidly ( $T_{1/2}$  = 30 s) than that formed with GABP $\beta$ 1. Because the  $\beta$ 1 and  $\beta$ 2 isoforms differ only at their COOH-termini, this part of the protein may be involved in stabilizing DNA binding.

The observations outlined thus far indicate that the  $\beta$ 1 and  $\beta$ 2 isoforms of GABP do not bind to DNA alone, but associate with the  $\alpha$  subunit to augment DNA binding. The question arises, therefore, as to whether the  $\beta$  subunits cause a conformational change in  $\alpha$  leading to its more avid interaction with DNA. Alternatively, or in addition

to causing a conformational change, the  $\beta$  subunits might, in association with GABP $\alpha$ , establish direct contact with DNA.

To determine whether GABP $\beta$ 1 contacted DNA when complexed with GABP $\alpha$ , DNA:protein complexes were exposed to ultraviolet (UV) light under conditions expected to permit covalent crosslinking between DNA and intimately bound proteins (L. A. Chodosh in Current Protocols in Molecular Biology, vol II, F. M. Ausubel et al., eds. (Greene/Wiley, New York, 1988). UV crosslinking was performed using an oligonucleotide composed of a GA binding site flanked by 10 bp of non-specific sequence ((5' AACCAAGCTTGCGGAACGGAAGCGGAAACCG 3') corresponding to residues located between 280 and 300 bp upstream of the herpes simplex virus gene encoding ICP4. Oligonucleotides were labeled to high specific activity by fill-in reaction with the Klenow fragment of DNA polymerase I in the presence of all four  $^{32}$ P-labeled dNTPs. DNA binding reactions were performed as described in a 96-well culture dish, followed by exposure to ultraviolet light. Samples were boiled in SDS-sample buffer and subjected to electrophoresis on SDS-polyacrylamide gels. Crosslinked protein species were visualized by autoradiography. The GABP subunits were incubated with  $^{32}$ P-labeled DNA that contained the purine-rich hexanucleotide repeats, exposed to UV light, and subject to electrophoresis on a denaturing polyacrylamide gel. When DNA was incubated with GABP $\alpha$  and exposed to UV light, a crosslinked product was observed bearing an electrophoretic mobility close to that of GABP $\alpha$  (Fig. 9). The appearance of this product was dependent on the presence of GABP $\alpha$  and increased in a time-dependent manner upon exposure to UV light. Moreover, it was eliminated by the inclusion of excess, unlabeled DNA that

contained the purine-rich hexanucleotide repeats, but not by excess non-specific DNA.

No evidence of protein:DNA crosslinking was observed when  $^{32}\text{P}$ -labeled DNA was mixed with GABP $\beta$ 1 and exposed to UV light. However, when the mixture of GABP $\alpha$  and GABP $\beta$ 1 was complexed with DNA and irradiated, new crosslinked products were observed. In addition to GABP $\alpha$ , two closely migrating polypeptide bands, slightly larger than the native size of GABP $\beta$ 1, became covalently attached to the radioactive DNA substrate (Fig. 9). Although GABP $\beta$ 1 was incapable of binding DNA on its own, when present in a ternary complex it appeared to be even more susceptible to UV-mediated crosslinking than GABP $\alpha$ . These data provide evidence that the  $\beta$ 1 subunit of GABP associates closely with DNA when complexed with GABP $\alpha$ .

#### Example 5

##### Formation of a Stable Complex between GABP $\alpha$ and GABP $\beta$ in the Absence of DNA

Having found that the  $\alpha$  and  $\beta$  subunits of GABP formed a heteromeric complex when exposed to their specific DNA substrate, gel filtration chromatography was used to determine whether these subunits might associate in the absence of DNA. Gel filtration chromatography was performed using a Superose-6 column (10 x 300 cm, Pharmacia) in buffer B supplemented with 0.4M NaCl. The column was calibrated with molecular weight standards thyroglobulin, apoferritin, catalase, bovine serum albumin and ribonuclease. 50-100  $\mu\text{g}$  of each protein was chromatographed at 0.5 ml/min. Elution volume was converted to  $K_{av}$  by the equation  $K_{av} = (V_e - V_0)/V_t - V_0$  where  $V_0$  = void volume = 8.1 ml;  $V_t$  = total bed

volume = 24.0 ml;  $V$  = eluted volume. The Stokes radius was calculated from a plot of  $(-\log K_{av})^{1/2}$  versus Stokes radius (Ackers, Adv. Prot. Chem. 24, 343 (1970)). GABP $\alpha$  eluted as a single peak at 15.2 ml; GABP $\beta$ 1 at 14.0 ml; GABP $\beta$ 2 at 15.8 ml. A mixture of equal amounts of GABP $\alpha$  and  $\beta$ 1 chromatographed as a single peak at 12.1 ml. The mixture of GABP $\alpha$  and GABP $\beta$ 2 chromatographed as a single peak at 13.9 ml. Loaded separately, GABP $\alpha$  and GABP $\beta$ 1 eluted as single peaks at  $K_{av}$  values of 0.45 and 0.38 respectively. However, when loaded together at an equimolar ratio, both subunits eluted as a single peak at  $K_{av}$  0.26 (Table 1). Analysis of column fractions by polyacrylamide gel electrophoresis confirmed that the peak at  $K_{av}$  0.26 contained both subunits.

TABLE 1

Determination of molecular weights of GABP subunits. Purified GABP $\alpha$ , GABP $\beta$ 1 and GABP $\beta$ 2 produced in *E. coli* were analyzed by gel filtration and sedimentation velocity as described (17,18).  $K_{av}$  values were calculated from the elution volume of a Superose-6 FPLC column. Apparent molecular weights were determined from  $K_{av}$  vs. log MW for the column. Stokes radii were determined from a plot of  $(-\log K_{av})^{1/2}$ . Sedimentation coefficients together with measured Stokes radii were used to calculate native molecular weights.

Protein	$K_{av}$	Apparent MW ( $K_{av}$ vs. log MW)	Stokes Radius	Sedimen- tation Coefficient	Corrected MW
GABP $\alpha$	0.449	158,000	49.3	3.1	66,000
GABP $\beta$ 1	0.377	349,000	63.0	3.1	82,000
GABP $\beta$ 2	0.486	106,000	44.0	2.5	46,000
GABP $\alpha$ + GABP $\beta$ 1	0.255	$1.3 \times 10^6$	87.7	4.5	170,000
GABP $\alpha$ + GABP $\beta$ 2	0.367	390,000	64.8	3.8	104,000

Evidence confirming the stable association of GABP subunits in the absence of DNA was also obtained from measurements of sedimentation velocity (Table 1). The gel filtration and sedimentation properties of a protein or protein complex are affected both by size and molecular shape. However, by using the analytical methods of Siegel and Monty (Martin et al., J. Biol. Chem. 236, 1372 (1961)), it was possible to calculate native molecular weights of the various protein species. Sedimentation coefficients were determined on 4.5 ml 10-30% glycerol gradients in 25 mM Tris-HCl pH 8.0, 75 mM NaCl, 0.75 mM EDTA, 1 mM DTT. 30 µg of each protein was loaded in 0.1 ml together with catalase, bovine serum albumin, and cytochrome c as internal standards. Gradients were centrifuged at 4°C for 40 hours at 39,000 rpm. Fractions (0.25 ml) were collected and analyzed by SDS-PAGE with Coomassie blue staining. The S value for each sample was determined by its sedimentation relative to the BSA and cytochrome c standards. Native molecular weights were derived using the Stokes radius together with measured sedimentation coefficients as described in Siegel et al, (Biochim. Biophys. Acta 112, 346 (1966)). Partial specific volume was calculated using the predicted amino acid sequences of each GABP subunit as described by Cohn et al, Proteins, Amino Acids and Peptides as Ions and Dipolar Ions, E. J. Cohn and J. T. Edsall, eds. (Reinhold Publishing Co., New York, 1943), pp. 370-381). As shown in Table 1, the calculated molecular weights of GABPα and GABPβ2 corresponded closely to their predicted sizes (51.3 and 37 kD, respectively). In contrast, GABPβ1 exhibited a native molecular weight (82 kD) roughly twice its expected size (41.3 kD).



The complex formed between GABP $\alpha$  and GABP $\beta$ 1 eluted from the gel filtration column prior to the largest molecular weight standard. Thus, the calculated molecular weight of this complex (170 kD) represents a provisional assignment. Since GABP $\beta$ 1 existed as a stable dimer on its own, the very large complex formed between it and GABP $\alpha$  was tentatively identified as a tetramer composed of two molecules of each subunit. Polyacrylamide gel analysis of the constituents of the GABP $\alpha$ :GABP $\beta$ 1 complex were consistent with this interpretation, showing that the subunits existed in equal stoichiometries. This interpretation was also consistent with the properties of the complex formed between GABP $\alpha$  and the  $\beta$  subunits was interpreted to reflect the dimer forming property of GABP $\beta$ 1.

Gel filtration and gradient sedimentation assays indicated that GABP $\beta$ 1 might exist as a dimer. This interpretation was tested using glutaraldehyde crosslinking assays. Bacterially expressed GABP $\beta$ 1 and GABP $\beta$ 2 were exposed to glutaraldehyde and subjected to electrophoresis on a denaturing polyacrylamide gel. Incubation of GABP $\beta$ 1 with glutaraldehyde led to the formation of a second polypeptide band exhibiting an apparent molecular weight roughly double that of the monomeric form of the protein (Fig. 10). Similar experiments conducted with GABP $\beta$ 2 failed to yield an analogous product.

Additional evidence suggesting that GABP $\beta$ 1 exists as a dimer resulted from crosslinking experiments with the intact polypeptide and a truncated form lacking the 110 NH<sub>2</sub>-terminal residues (BN110, see Fig. 12B). Crosslinking of the truncated protein led to the formation of an additional species roughly double the size of the monomeric form. When the truncated protein was

mixed with intact GABPB1 and exposed to glutaraldehyde, three crosslinked protein species were observed. Two species corresponded to crosslinked, homodimeric complexes that had been  
5 observed upon glutaraldehyde treatment of active GABPB1 and the NH<sub>2</sub>-terminal truncated derivative. The third species migrated between the presumed homodimeric forms and probably represented a  
10 heteromeric complex consisting of one GABPB1 polypeptide and one truncated polypeptide. Therefore, both molecular weight measurements and crosslinking assays showed that GABPB1 but not GABPB2 exists as a stable homodimer.

#### Example 6

#### 15 Mapping of Functional Domains of GABP $\alpha$ and GABPB

Experimental results described above indicate that GABP $\alpha$  should contain at least two functional components, one that facilitates DNA binding and another that allows complex formation  
20 with GABPB. The GABPB1 polypeptide should contain at least three components, facilitating self-dimerization, heterodimerization with GABP $\alpha$ , and direct contact with some part of the purine-rich DNA substrate. Recombinant copies of the genes that  
25 encoded each subunit were systematically deleted to localize these components. Deletion mutants of GABP $\alpha$  were generated by polymerase chain reaction and expressed in pT5 as described (Breedon et al, Nature 329, 651 (1987)). Soluble bacterial extracts  
30 containing deleted variants of GABP $\alpha$  were used for binding reactions. NH<sub>2</sub>-terminal deletions of GABPB1 were generated by exonuclease III digestion, followed by digestion with S1 nuclease and ligation of Bam HI linkers. All deletions were sequenced and

subcloned into the appropriate pET3 vector (Rosenberg et al., Gene 56, 125 (1987)) to maintain the proper reading frame. COOH-terminal deletions were generated using 3' deletions of the cDNA inserted in Bluescript (Stratagene) by subcloning Bst EII-Asp 718 or Sac I-Asp 718 fragments into the pET-GABP $\beta$ 1 plasmid that had been digested with the appropriate enzymes. Translation termination codons were provided by vector sequences so that in some cases extra amino acids are appended to the open reading frame. All GABP $\beta$  derivatives were insoluble and re-solubilized in 8M urea followed by dialysis against 10 mM Tris pH 8.0, 75 mM KCl or NaCl, 1 mM DTT, 0.2mM PMSF, 1mM benzamidine, 10% glycerol prior to use in binding reactions. All derivatives were expressed equivalently as determined by Coomassie staining of SDS polyacrylamide gels.

Deletion variants of GABP $\alpha$  that were missing as many as 313 residues from the NH<sub>2</sub>-terminus retained the capacities to bind DNA and complex with GABP $\beta$  (Fig. 11). A GABP $\alpha$  variant further missing 17 residues ( $\alpha$ N313/437) from the COOH-terminus also retained both functions. More extensive deletion from the COOH-terminus, to amino acid 407 ( $\alpha$ N313/C407), yielded a protein that was capable of binding to DNA, but had lost the ability to complex with GABP $\beta$ 1. These results showed that the Ets-related domain of GABP $\alpha$  was sufficient for DNA binding. The region of GABP $\alpha$  required to form a complex with GABP $\beta$  included the Ets-related segment, as well as 37 amino acids located on the immediate COOH-terminal side of the Ets-related domain.

In order to define regions of GABP $\beta$ 1 that interact with GABP $\alpha$  and contact DNA, systematically deleted variants were produced and tested in gel retardation and UV-crosslinking assays (Fig. 12). Variants that lacked up to 228 residues ( $\beta$ C154) from

the COOH-terminus of GABPB1 proved to be functional in both assays. Deletion of an additional 33 residues (BC121) yielded a protein that failed to function in either assay. The boundary defined by these experiments corresponded to the location of the most COOH-terminal of four 33-amino acid repeats that are present in both isoforms of GABPB.

Although about 70% of GABPB1 could be deleted from its COOH-terminus without eliminating interaction with GABP $\alpha$  and DNA, removal of only a small segment from the NH<sub>2</sub>-terminus resulted in deleterious effects. A variant of GABPB1 that lacked 19 NH<sub>2</sub>-terminal residues (BN19) was slightly less effective in converting GABP $\alpha$ -derived complexes into the very slowly migrating heteromeric complex. When tested in the UV crosslinking assay, BN19 yielded a reduced amount of crosslinked product relative to the intact B1 isoform. Variants that lacked 47 and 67 residues (BN47 and BN67) were progressively more defective in the complex formation assay and failed to be crosslinked to the radioactive DNA probe as efficiently as the intact protein. Finally, variants missing 80 or more residues from the NH<sub>2</sub>-terminus were completely defective in both assays. The progressive loss of function observed in deleted forms of GABPB1 corresponded to the progressive loss of the 33-amino acid repeats (Fig. 12). BN19 was truncated within the first of the repeats, BN47 within the second, BN67 after the second, and BN80 within the third repeat. The functional properties of the GABPB1 deletion mutants indicate that the 33-amino acid repeats are important both for complex formation with GABP $\alpha$  and DNA contact.

Example 7

## A Model for DNA-Bound GABP

The foregoing observations have been incorporated into a provisional model of the complex formed when GABP $\alpha$  and GABP $\beta$ 1 associate with a directly repeated set of purine-rich hexanucleotides (Fig. 13). Each hexanucleotide repeat is hypothesized to be contacted by both GABP $\alpha$  and GABP $\beta$ 1. The linear order of contact, wherein GABP $\alpha$  is associated with guanines on one side of each hexanucleotide and GABP $\beta$  with adenines on the other side, was deduced from three separate observations. First, GABP $\alpha$  was alone capable of protecting both guanines from methylation by dimethylsulfate. Second, the DNase I footprint generated by the mixed subunits, relative to that resulting from GABP $\alpha$  alone, was extended in a direction toward the adenine residues of the hexanucleotide repeat. Third, addition of GABP $\beta$  caused an enhanced pattern of methylation of adenine residues relative to the pattern generated in binding reactions that contained GABP $\alpha$  alone.

The most stable complexes formed between GABP and DNA were observed with the mixture of GABP $\alpha$  and GABP $\beta$ 1. The  $\beta$ 1 subunit, unlike  $\beta$ 2, was observed to exist as a stable homodimer. Moreover, when mixed with GABP $\alpha$ , the  $\beta$ 1 subunit generated a high molecular weight complex probably consisting of two polypeptides of each subunit. This heteromeric tetramer is believed to bind in a concerted manner to two purine-rich hexanucleotide repeats. If that is the case, a flexible region should exist between the dimerization domain of GABP $\beta$ 1 and the surfaces located near its NH<sub>2</sub>-terminus that facilitate interaction with GABP $\alpha$  and DNA. Without such

flexibility a linked set of polypeptides would not likely be capable of binding simultaneously to a DNA substrate that is not rotationally symmetric.

#### Example 8

#### 5            Isolation of cDNA Clones Encoding Human GABP                  Alpha and Beta Subunits

                 cDNA clones encoding human GABP alpha, beta3 and beta4 were isolated by screening a human fetal brain cDNA library. cDNAs for human GABP  
10            beta1 and beta2 were isolated from a HeLa cell cDNA library. The probes used to screen for human GABP alpha were the 865 bp Ava 1-Sst 1 and 678 bp Bam H1-Sst 1 fragments of the mouse GABP alpha cDNA. The probe used to isolate the human beta was an 850 base  
15            pair fragment from the 5' end of the mouse GABP beta2 cDNA.

                 Purified DNA fragments used as probes were radiolabeled with  $^{32}\text{P}$  by random priming reactions. Hybridization conditions were 6X SSC, 1X Denhardt's,  
20            0.05% sodium pyrophosphate, 100  $\mu\text{g/ml}$  yeast tRNA. The final wash buffer was 2 X SSC. The hybridization and washing temperature were 65° C.

                 cDNAs were confirmed as the human homologs of mouse GABP by determination of their nucleotide  
25            sequence.

                 Six different cDNAs have been deposited with the American Type Culture Collection (ATCC), 12301 Parklawn Drive, Rockville, MD 20852.

                 1A = human GABP $\beta$ , Eco R1 fragment common  
30            to all beta isoforms, in Bluescript KS+, isolated from HeLa cDNA library.

                 4A = human GABP $\beta$ 1 Eco R1 fragment of Bluescript KS+, isolated from HeLa cDNA library.

5A = human GABPB2, Eco R1 fragment in  
Bluescript KS+, isolated from HeLa cDNA library.

5 F = human GABPB3, Eco R1-Xho 1 fragment in  
Bluescript SK-, isolated from human fetal brain cDNA  
library.

J = human GABPB4, Eco R1- Xho 1 fragment  
in Bluescript SK-, isolated from human fetal brain  
cDNA library.

10 G = human GABP $\alpha$ , Eco R1- Xho 1 fragment in  
Bluescript SK-, from human fetal brain cDNA library.

\* \* \* \*

The entire contents of all documents cited  
herein are hereby incorporated by reference.

15 While various aspects of the invention  
have been described in some detail for purposes of  
clarity and understanding, one skilled in the art,  
from a reading of this disclosure will appreciate  
that various changes can be made in form and detail  
without departing from the true scope of the  
20 invention. One skilled in the art will also  
appreciate that the invention includes the human  
counterparts of the sequences specifically disclosed  
herein as well as the encoded amino acid sequences,  
and fragments thereof. The invention also relates  
25 to the deposited human sequences and fragments  
thereof as well as to the full length sequences of  
which the partial sequences form a part.

5     WHAT IS CLAIMED IS:

1.     A DNA segment encoding a subunit of  
GA binding protein (GABP), or an epitope specific  
10    thereto, or a DNA fragment complementary to said DNA  
segment.

15     2.     The DNA segment according to claim 1,  
wherein said GABP is human GABP.

3.     The DNA segment according to claim 1  
20    wherein said subunit is GABP $\alpha$ .

4.     The DNA segment according to claim 1  
25    wherein said subunit is GABPB1.

5.     The DNA segment according to claim 1  
wherein said subunit is GABPB2.

30     6.     The DNA segment according to claim 3  
wherein said subunit has the amino acid sequence  
shown in Figure 2A.

35     7.     The DNA segment according to claim 4  
wherein subunit has the amino acid sequence shown in  
Figure 2B.

40     8.     The DNA segment according to claim 5  
wherein said subunit has the amino acid sequence  
shown in Figure 2B.

45     9.     A recombinant DNA molecule  
comprising:

- i)    said DNA segment according to claim 1;
- and
- ii)   a vector.



5           10. A host cell stably transformed with  
said recombinant DNA molecule according to claim 9.

10           11. The host cell according to claim 10  
wherein said cell is a procaryotic cell.

15           12. The host cell according to claim 10  
wherein said cell is a eucaryotic cell.

20           13. A method of producing a recombinant  
GABP subunit protein, or portion thereof defining at  
least an epitope specific thereto, comprising  
culturing said host cell according to claim 10 under  
conditions such that said segment is expressed and  
25 said protein thereby produced, and isolating said  
protein.

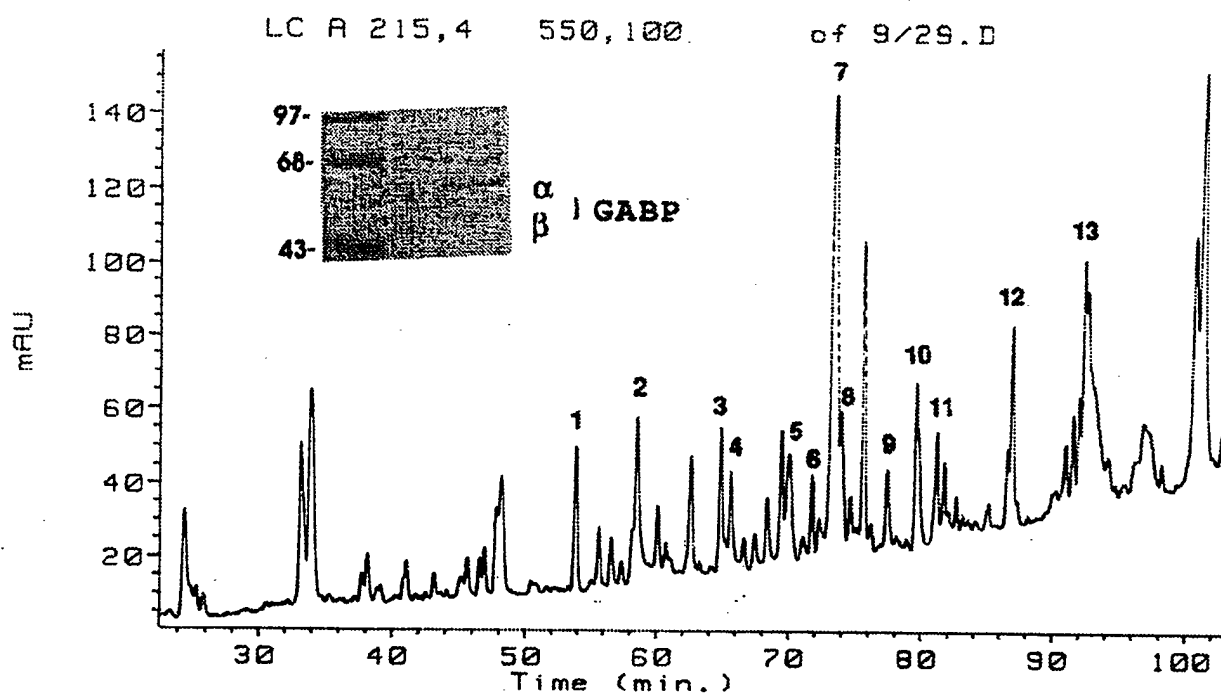


FIG. 1



## FIG. 2A2

841	CACGCGGAAGCAGAAGCGCATCTCGTTGAAGAAGCTCAAGTGATAACTCTTGACGGCACC H A E A E A H L V E E A Q V I T L D G T	149
901	AAGCACATTACGACCATTTTCAGACGAGACCTCGGAGCAGGTGACGAGATGGGCTGCTGCA K H I T T I S D E T S E Q V T R W A A A	169
961	CTGGAAGGCTACAGAAAGACGAGAGCGCCTTGCGCATCCCCCTATGATCCTATACGCTGG L E G Y R K E Q E R L G I P Y D P I R W	189
1021	TCCACGGACCAAGTCCTGCATTTGGGTGGTTTGGGTAATGAAGGAGTTTCAGCATGACTGAT S T D Q V L H W V V W V M K E F S M T D	209
1081	ATAGACCTCACCCACTCAACATTTTCGGGAAGAGAATTATGTAGTCTCAACCAAGAAGAT I D L T T L N I S G R <u>E L C S L N Q E D</u>	229
1141	TTTTTTCAGCGGGTCCCTCGGGGAGAAATTCTTTGGAGTCATCTGGAGCTTCTTCGAAAA <u>F F Q R V P R G E I L W S H L E L L R K</u>	249
1201	TATGTTTTCGGCCAGCCCAAGAGCAACAGATGAATGAGATAGTTACCATTTGACCGCCTGTG <u>Y V L A S Q E Q Q M N E I V T I D Q P V</u>	269
1261	CAGATTATTCAGCCTCAGTGCCCTCCGGCTACACCGACTACGATTAAAGTTATAACAGC Q I I P A S V P P A T P T T I K V I N S	289
1321	AGTGCAAAAGCAGCTAAAGTGCAACGGTCCCAAGGATTTTCAGGAGAAGACAGAAGTTCA S A K A A K V Q R S P R I S G E D R S S	309
1381	CCGGGGAACAGAAACAATGGTCAGATCCAACTATGGCAGTTTTCCTAGAACTT <u>P G N R T G N N G Q I Q L W Q F L L E L</u>	329

## FIG. 2A3

1441 CTTACTGACAAGGATGCTCGAGACTGTATTTCTTGGTGGTGATGAAGTGAATTTAAG  
L T D K D A R D C I S W V G D E G E F K 349

1501 CTAAATCAGCCTGAGTTGGTTGCGCAAAAATGGGGACAACGTAAGAACCAAGCCTACCATG  
L N Q P E L V A Q K K W G Q R K N K P T M 369

1561 AACTATGAGAAACTTAGCCGTGCATTACGGTATTATTATGATGGGGACATGATTGTAAA  
N Y E K L S R A L R Y Y Y D G D M I C K 389

1621 GTTCAAGGCAAGAGATTTGTGTACAAATTTGTTGTGACTTGAAGACTCTTATTGGATAC  
V Q G K R F V Y K F V C D L K T L I G Y 409

1681 AGTGCAGCAGAACTGAACCGTCTGGTCAATAGAGTGTGAACAGAGAACTGGCAGCGGATG  
S A A E L N R L V I E C E Q K K L A R M 429

1741 CAGCTGCATGGGATTGCCAGCCAGTCACGGCAGTAGCAGTGGCAGCCACCTCTCTACAG  
Q L H G I A Q P V T A V A L A A T S L Q 449

1801 GCAGACAAAAGAGATTTGAGACCTAGGACCTCCTGGGGAGTCTTAAAGGTTTTTCTTAAATA  
A D K E I 454

1861 TTTAGAGCAAGCTTTGCTCTAACCTTTATTACTGAATTTGAATCGTATTTCTAGAGTGTA  
1921 CAATCTGATGCATGATTTTTTTTATAAATAATTTTCATCTCTGTGAAAAA  
1981 AAAAA 1985

## FIG. 2B1

GABPβ1

1	GGAAATAGCGCCCTGTGCAGCGAAGCGCTCTGTGTGTGGCGGGCGCTCTGCCCTGA	16
61	CGGCTCCGAGGCGCGGTCCCGCACCTCCTCCCGTGTCTCCCGCGCGCTCCCG	
121	AAGCTTTTCCAGATGTCCCTGGTAGATTGGGAAGAAGCTTTTAGAAGCGGCACGAGCC	
	M S L V D L G K K L L E A A R A	16
181	GGTCAAGATGATGAAGTTCCGCAATTTTGATGGCAATGGAGCTCCTTTTACTACAGACTGG	
	G Q D D E V R <u>I L M A N G A P F T T D W</u>	36
241	TTGGGAACCTTCTCCACTTCATCTGGCCGCACAGTATGGGCATTTCTCTACCACAGAGGTT	
	L G T S P L H L A A Q Y G H F S T T E V	56
301	CTTCTCCGAGCCGGTGAAGTAGAGATGCCAGGACCAAGTGGACCGGACACCACTGCAC	
	L L R A G V S R D A R T K V D R T <u>P L H</u>	76
361	ATGGCGGCTTCTGAGGGCCATGCCAACATAGTAGAAGTTTGTGTTAAGCATGGTGTCTGAC	
	<u>M A A S E G H A</u> N I V E V L L K H G A D	96
421	GTCAATGCCAAAGGATATGTTAAAGATGACAGCTCTGCATTTGGCAACAGAACATAATCAT	
	V N A K D M L K M T A L H W A T E H N H	116
481	CAAGAGTGGTGGAGCTTTTAATCAAAATATGGTGTGATGTACACCGCAGAGTAAATTT	
	Q E V V E L L I K Y G A D V H T Q S K F	136
541	TGTAAACTGCATTTGATATTTCAATAGACAATGGAATGAAGATTTAGCAGAGATATTA	
	C K T A F D I S I D N G N E D L A E I L	156
601	CAGATTGCTATGCAGAACCAAAATCAACACCAACCCGGAGAGTCTGTGACACTGTGACAAATA	
	Q I A M Q N Q I N T N P E S P D T V T I	176

## FIG. 2B2

661 CACGCTGCCACACACAGTTCATCATTTGGACCCGGAGGGTGGTGAACCTCACAGATGAA 196  
H A A T P Q F I I G P G G V V N L T D E

721 ACAGGAGTATCTGCTGTTTCAGTTTGGAACTCCTCTACGTCAGTATTAGCTACGTTAGCT 216  
T G V S A V Q F G N S S T S V L A T L A

781 GCCTTAGCTGAAGCTTCTGCCCCCATTTGTCCAATTCTTCAGAAACTCCAGTGGCCACAGAG 236  
A L A E A S A P L S N S S E T P V A T E

841 GAAGTGGTTACCGCAGAAATCTGTGGATGGTGCAATTCAGCAAGTAGTTAGCTCAGGGGT 256  
E V V T A E S V D G A I Q Q V V S S G G

901 CAGCAAGTCATCACGATAGTTACAGATGGAATCCAGCTGGGGAATTGCACTCCATACCA 276  
Q Q V I T I V T D G I Q L G N L H S I P

961 ACCAGTGGGATGGGTCAGCCCCATCATTTGTGACGATGCCCGATGGACAGCAAGTATTGACA 296  
T S G M G Q P I I V T M P D G Q Q V L T

1021 GTACCAGCAACAGACATTGCTGAAGAACTGTTCATCAGTGAAGAGCCACCGCTAAGAGA 316  
V P A T D I A E E T V I S E E P P A K R

1081 CAGTGTATGGAATAATTGAGAGCCGGGTGGAATGTGCAGAAATTGAAGAGAGAGAAGCG 336  
Q C M E I I E S R V E C A E I E E R E A

1141 CTTCAGAAACAGCTGGATGAAGCCACCGAGAGGCCCGAGAAATACCGACAGCAGCTGCTT 356  
L Q K Q L D E A N R E A Q K Y R Q Q L L

## FIG. 2B3

1201 AAGAAGGAGCAGGCGCTACAGGCAGAAGCTGGAGGCCATGACACGCATCCAG  
K\_K\_E\_Q\_E\_A\_E\_A\_Y\_R\_Q\_K\_L\_E\_A\_M\_T\_R\_I\_Q 376

1261 ACCAACAAAGAAGCCGTTTAGCTGCCCATGAACACCAGTTTGTCTTTTACCTTTTGTCCAGA  
T\_N\_K\_E\_A\_Y 382

1321 AAGAATATAGTCTTGAACGTGTACACAGTAAGGACACAGCCATGGGAATACCGAATAATAG  
1381 AAAATACTACAGCTTGATAACGGGACTTAAGCCATGAGCTCTCAATTCCTTGTAAATATAA  
1441 ACTTTAGAACTTGTGAAATGTATTTAAAACTGAATTTCTGTAAATAGTTTTTGTGTTT  
1501 CTTTTTTTACTGTTATAAATGAGTTGATGAAGCTTGTGCTGAGATCCAGAGGCCACAGC  
1561 AAGCCACTGTACGGGAAATTCCTTTTGAATTCGGTGCAGACTTAATTTCTCAGAAACAG  
1621 AAGTTGTAAGCGTGTGTGCTCTTAGACCAAATGTGGAACAATATATGTTGGACTGAT  
1681 GCTGGAAATACTCTGCAAGAGATTTGCCCCAGGAACCTTTTGTACAGCTTTTAAGTTGTC  
1741 TCAGGTTCTCTGAAACATTTTTTAGGAAGCAGAAATTTTATAATTTGTTCTATTTTCAGCT  
1801 ACATGCAAGTAGATTTACATGTATATGAAGCAAAATTTTTTAAATTTTGTGTACATA  
1861 TCCTGCATGTTTATGATTTTAAACACATCCCTTGCTTAGGTATGTCTCTACAAAGAC  
1921 GATGAAAGTTACCAAGAAAAAACCAGGCACCACTGTCGTCTGCCGTCGTTGAGAGT  
1981 GAGTAAGCTCACAGCTGGTTTTTACTGGCGACCATGACAAATGTTCTTGGAGCAGCTTGCAA  
2041 GAGAACTCGAATCTGCAAAATATATGCTTTTCTGACAAATGTATTTGTGAGGATGTGAGAC  
2101 TGCTCTGTGCCCTTCCCTCCATCATGACTTCCACACGAGCACCTTTGAGGCGCTGGATTTTA  
2161 AGATAATGTTTTTGGCAAACTCAGTGCAATTTGGGTTTCTGAAATATTTTCATGGACTTACT  
2221 CCCCTCCCCTAAGAAATTAATTTCTTATGGAAATTTGCTTTTGTATGACAGCAGGAC  
2281 CTTTTGTACTCAGTGAGGCTATGGGTGTGCTAACGTAACCTTTCGGTATTTCCCTGTGAAC  
2341 GCGCTCCGGGCTGTGCTGCGACTCTCTCACAATATTTGTCATTTCTGAAATGTATGGACA  
2401 TTTGACAGACTATTCCATTCTCTTTTGTAAATGTAACCTCAAGCTCTTTATATTTTGT  
2461 TTTGTTCTATTTGCTTTATGAAAACGTTTGGCCCCCTCCCCACCCAGGCTCCTCTAGCTTA  
2521 TCTGTATAAAGAACCAAGTTACCTGCGAGACTGTACTCTGGCTCATTTTCTATTTTACAA  
2581 TTAAATAACTTTCCACATAAAAAAATAAAAAA 2616



## FIG. 2B4

GABP $\beta$ 2

1131	GTAAGGAGCCTTATACCCGGTGTGTTTGTGCTGCAGCCCATCCAAAATAAAATTTGTTTCAT	349
	V R S L I P G V F C C S H P K	
1191	TATTATTATTATCACTATTATTATTAATTTGTTGTTTATATTTACTGGCCAT	
1251	GTTTGTGATAAAGAAATGAAAGTGAATAGTGTCATCATTTTGTGTTTGGCCCTTAAACATATGG	
1311	CAAGCCCTCAATAAAATAAATATTGAATGAAAAAATAAAAAA 1356	

FIG. 3B

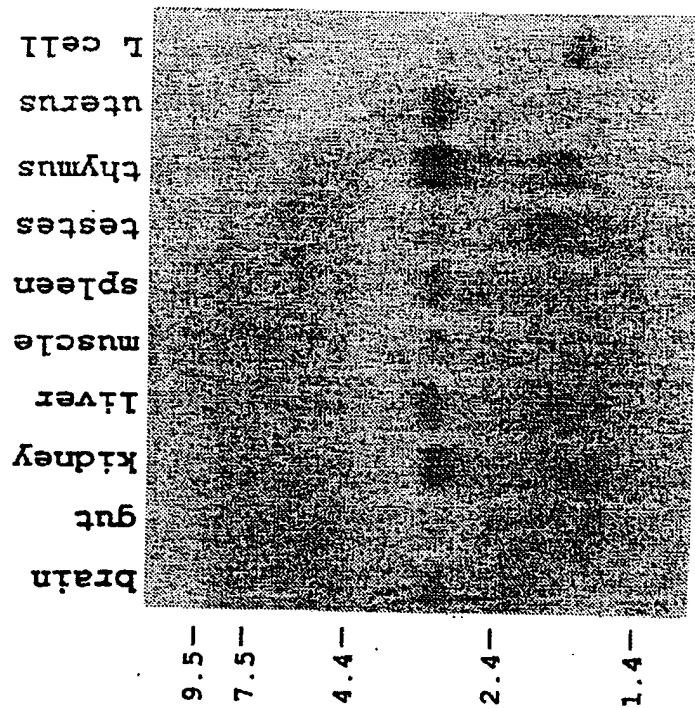


FIG. 3A

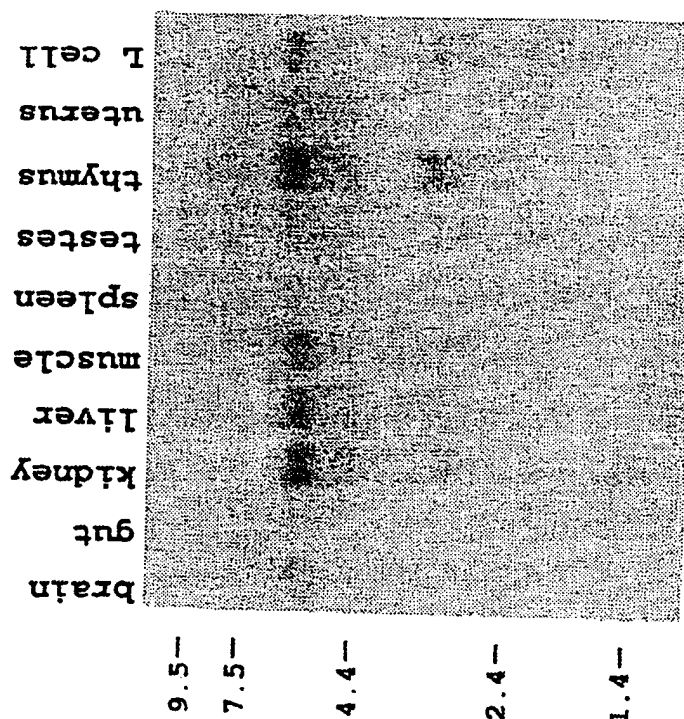


FIG. 4A

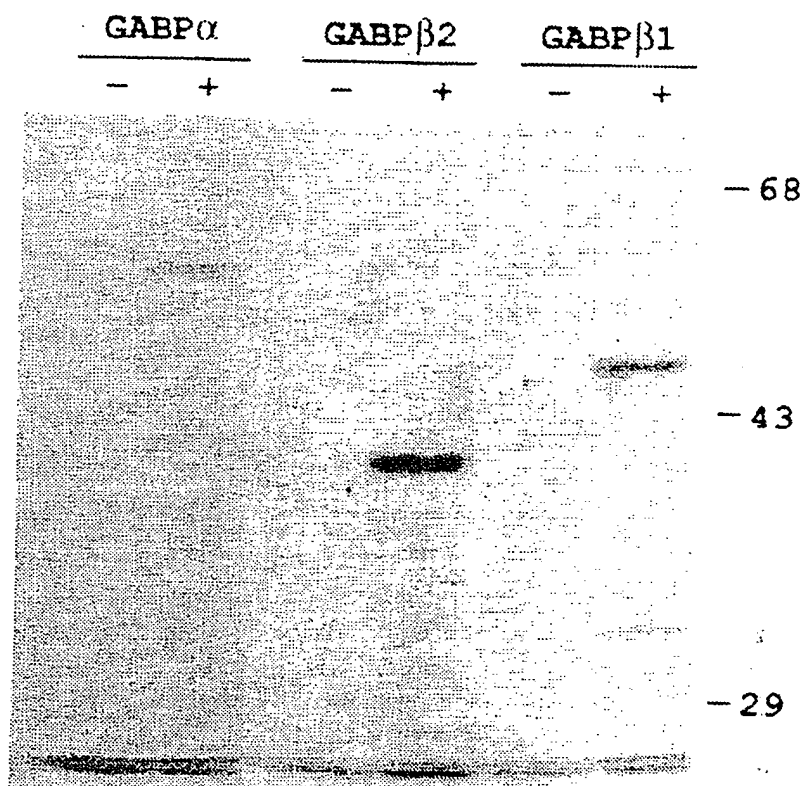


FIG. 4B

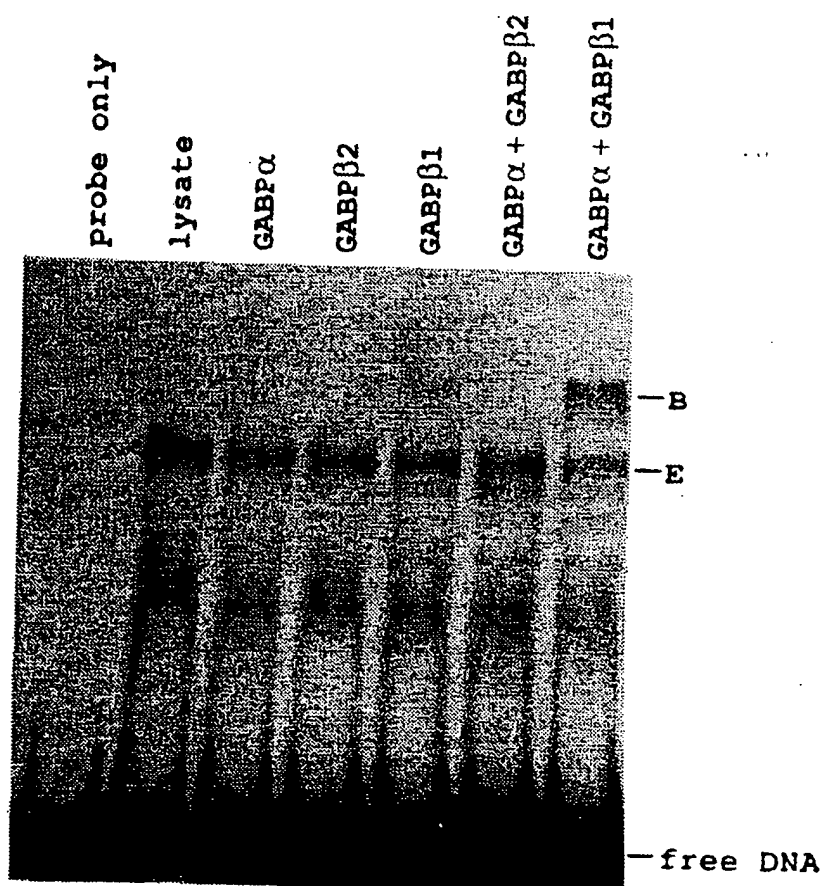
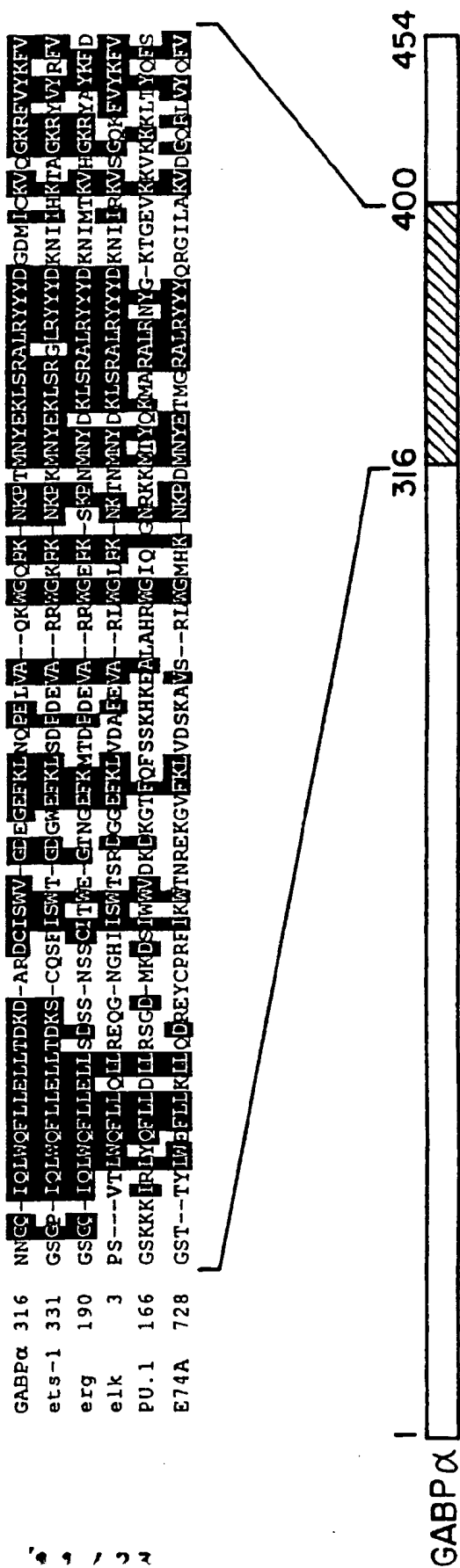


FIG. 5



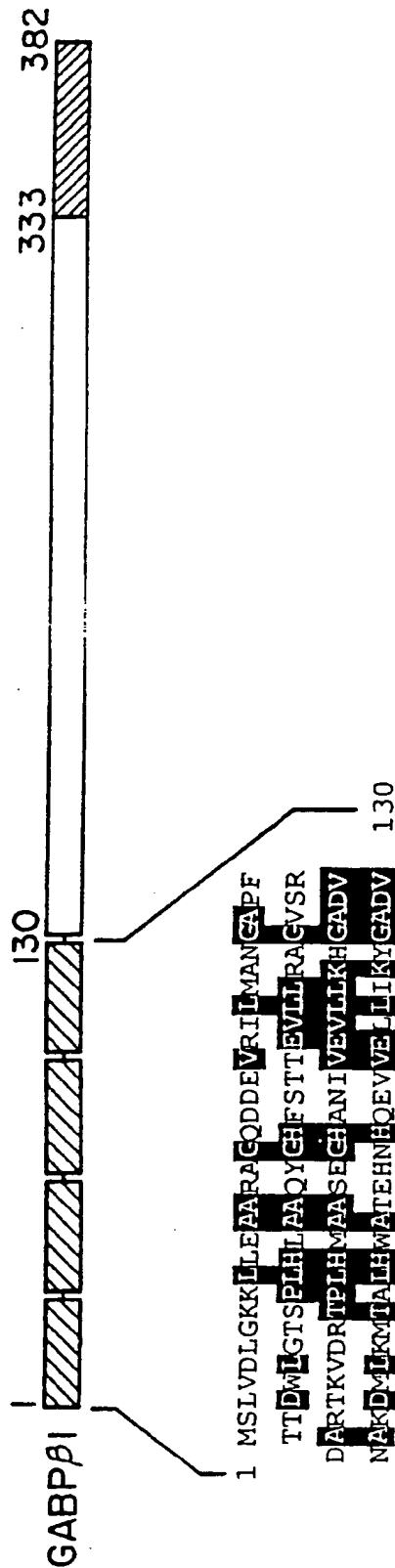


FIG. 5 con't.

GABPβ	-A-D-L---TPLH-AA---H---VEVLLK-GADV
cdcl0/SWI4,6	N-QD--G-TPLHWAA--AN--LV--L---GANL
Notch	NA-D--G-TPLHLAAR-----LL---A--
glp1	---D--GRTALHLAA-----V--L-----
lin12	---D--GRTALH-AA-N-----V-YL---A--
ankyrin	-D-----G-TPLH-AA--GH---V--LL--GA--
NFKB	---D--G-T-LHLA-----L---LL--GAD-
fem1	-----G-TPL--AA--GH---VVKLLE-G-D-
bcl3	-A-N--G-TPLH-AV-----L--VRLL--GA--
overall	
consensus:	---D--G-TPLH-AA-----V--LL--GA--

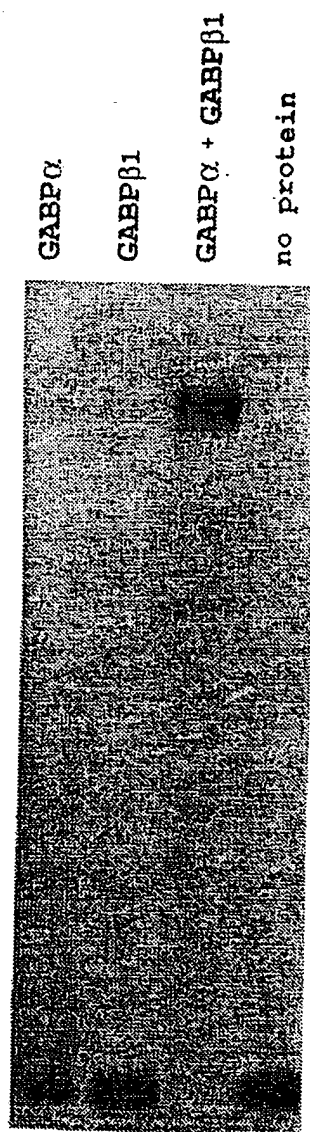


FIG. 6A

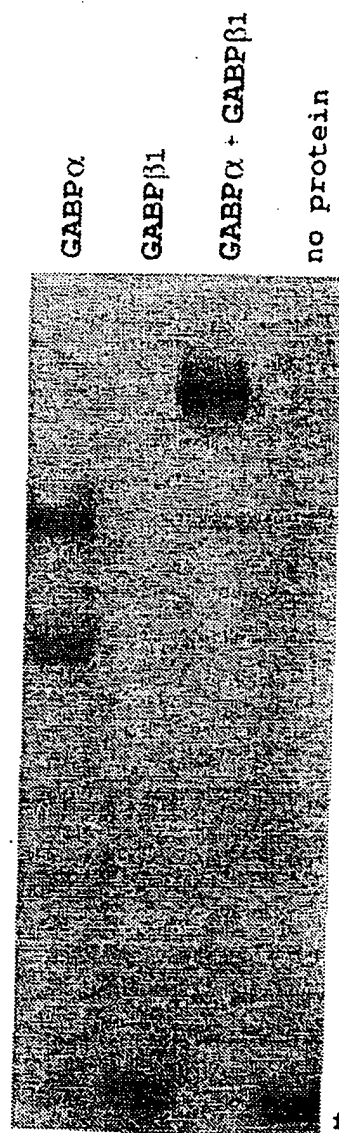


FIG. 6B

FIG. 7A1

FIG. 7A2

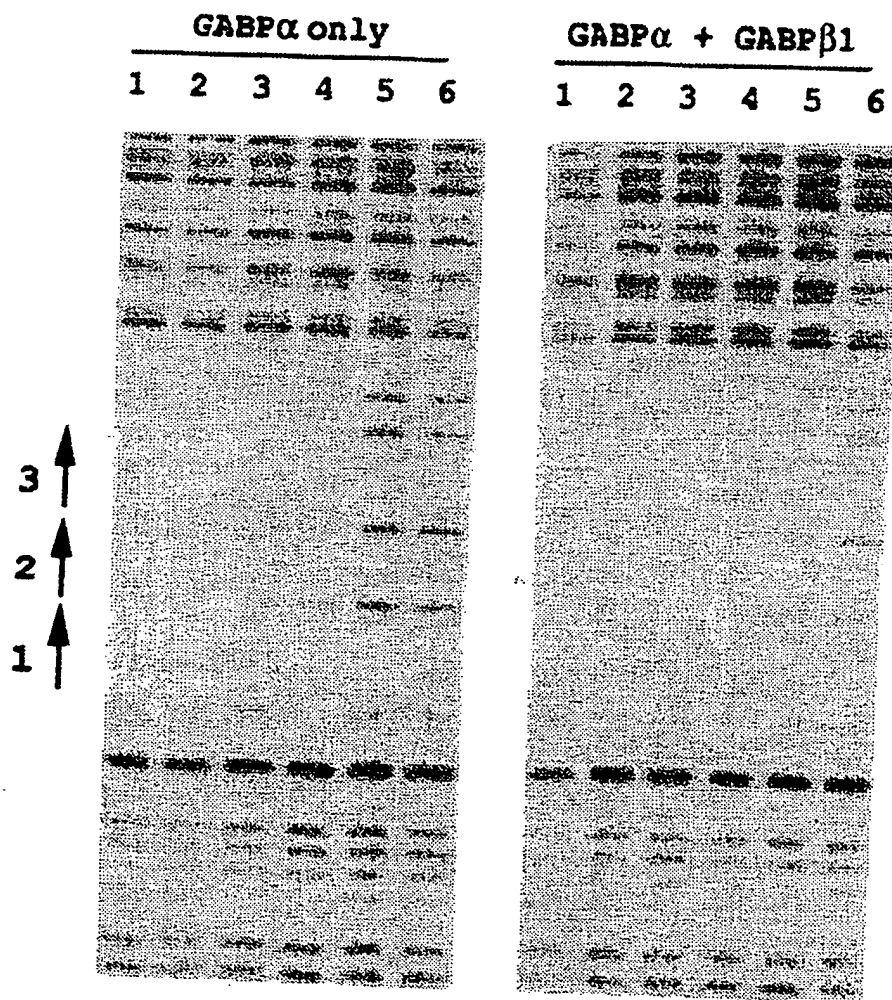
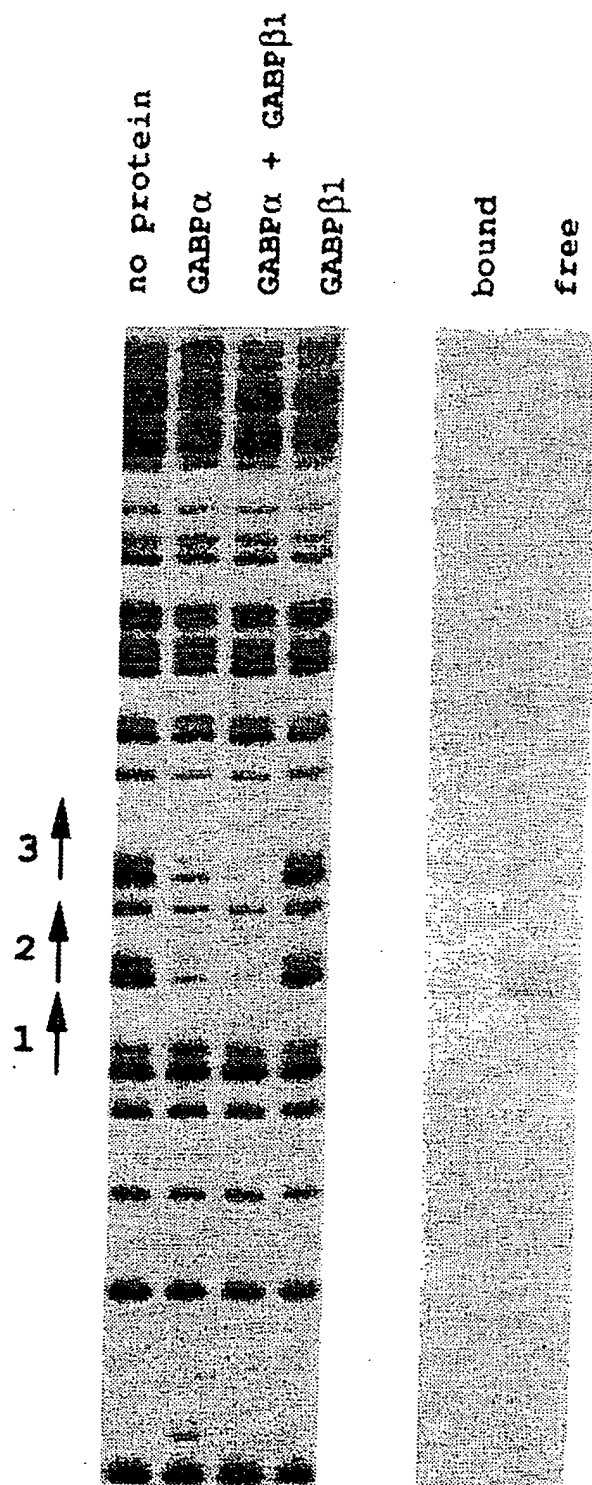


FIG. 7B1

FIG. 7B2



●●●●●    ●●●●●  
 TGC GGAAC **GGA** AGC GGAACCG  
 1      2      3



FIG. 8

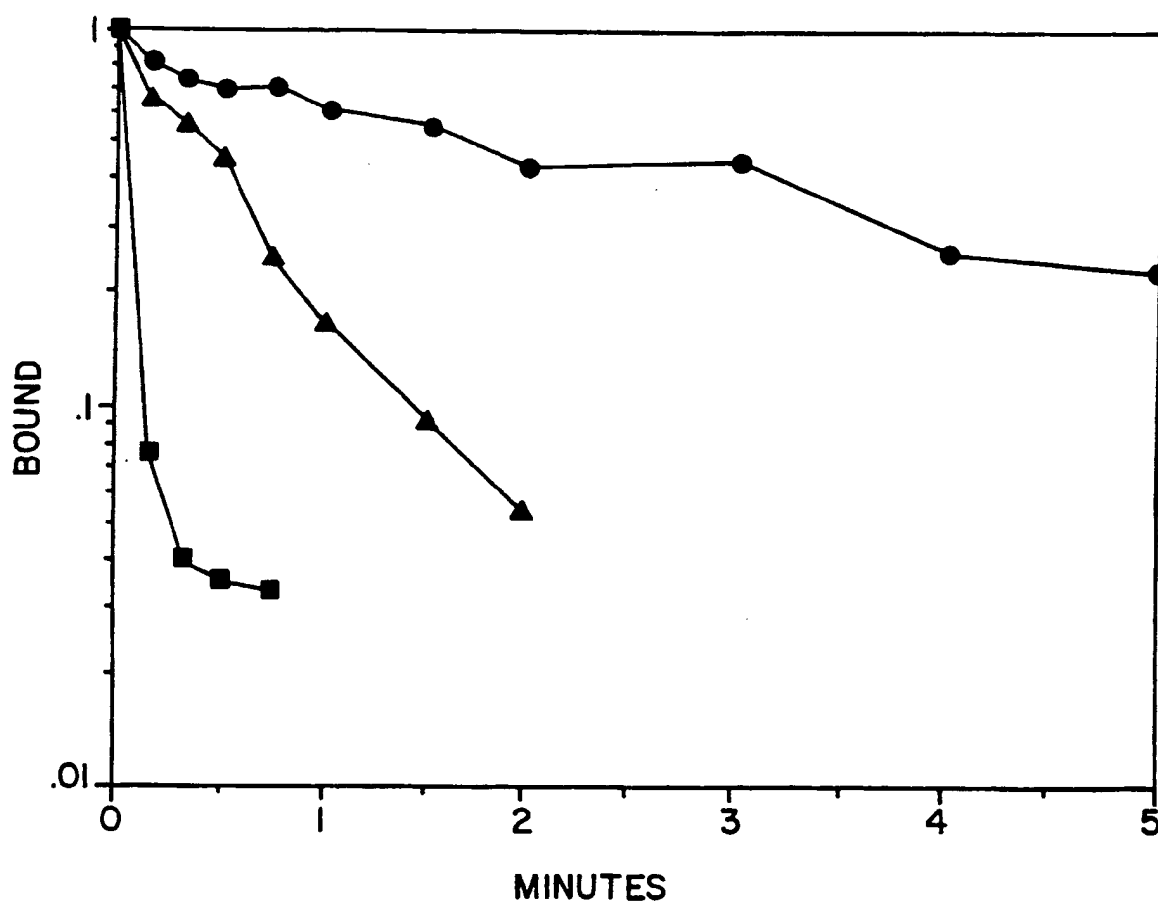


FIG. 9

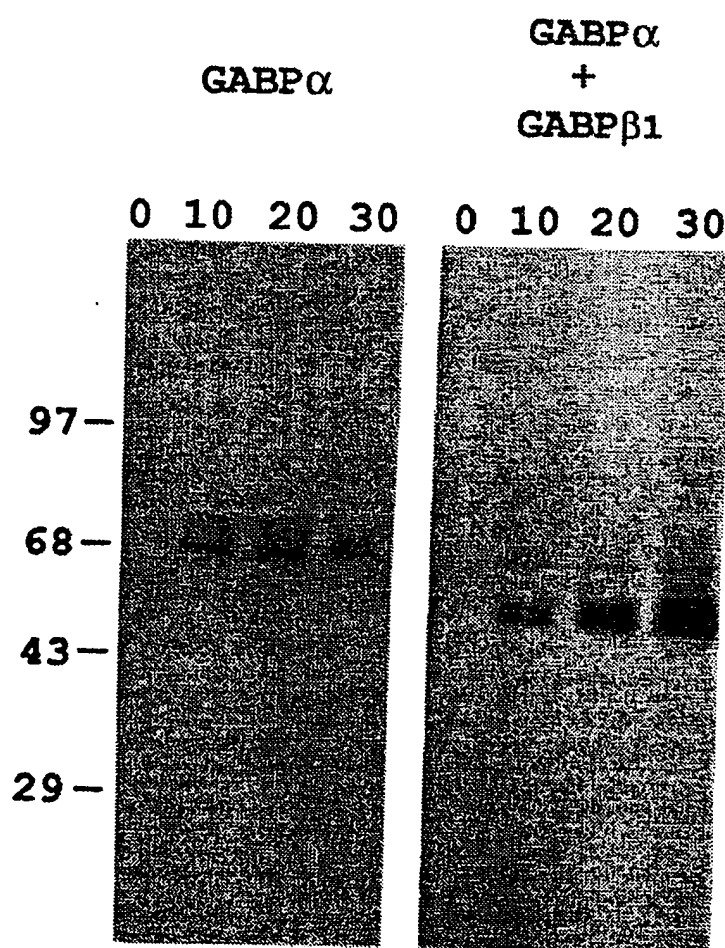


FIG. 10

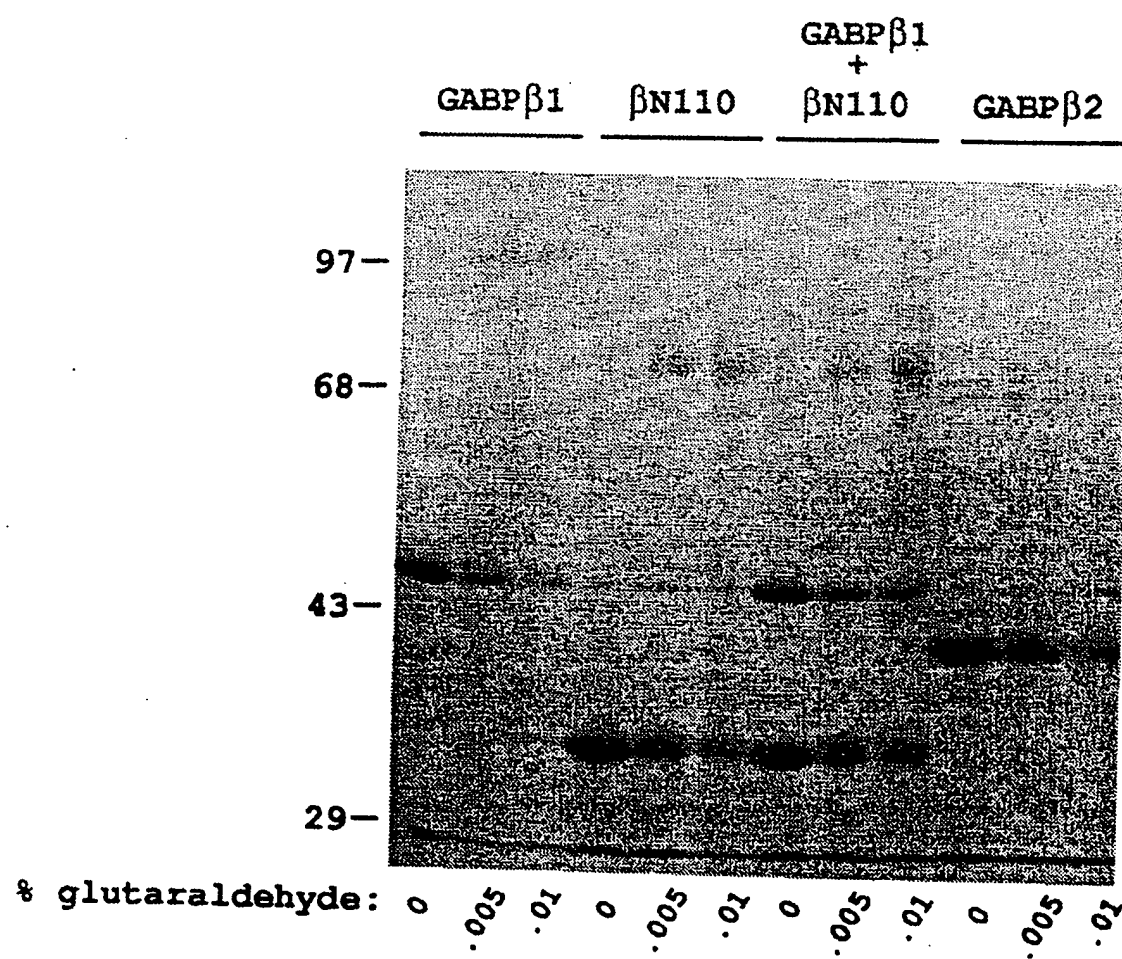


FIG. IIA

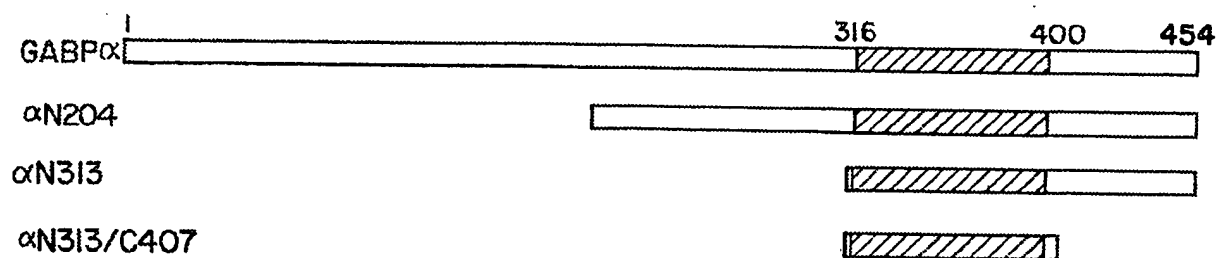
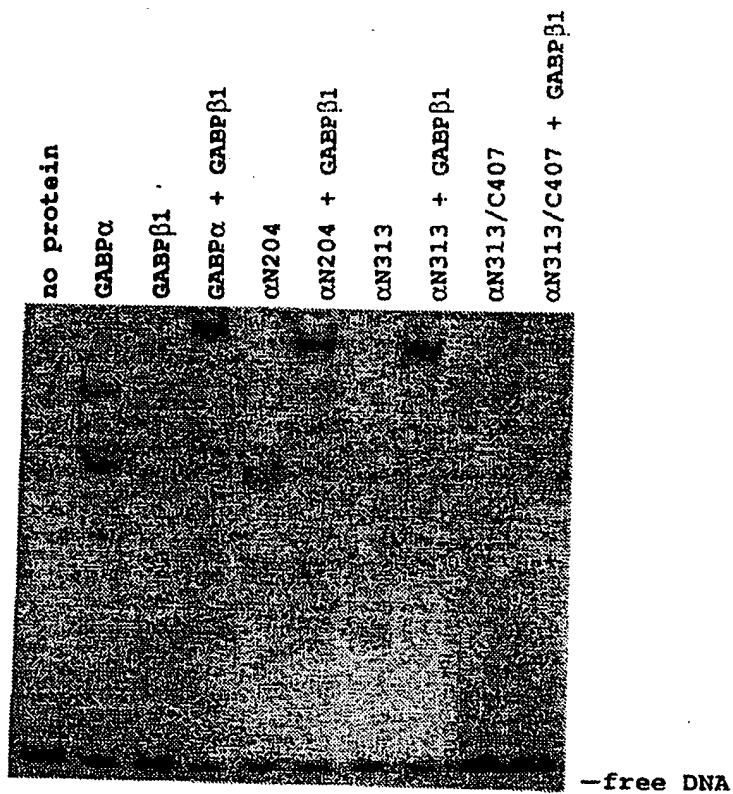


FIG. IIB



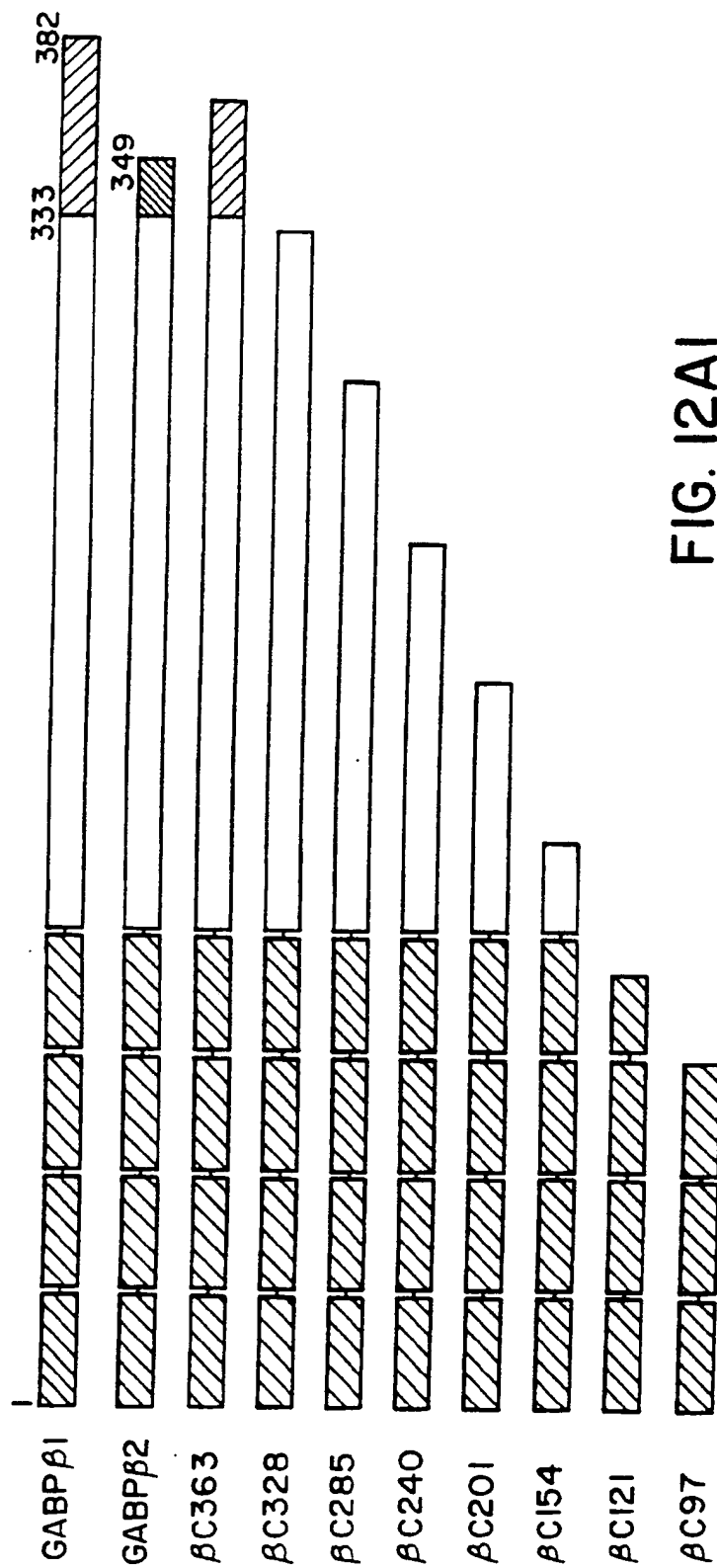


FIG. 12A1

FIG. 12A3

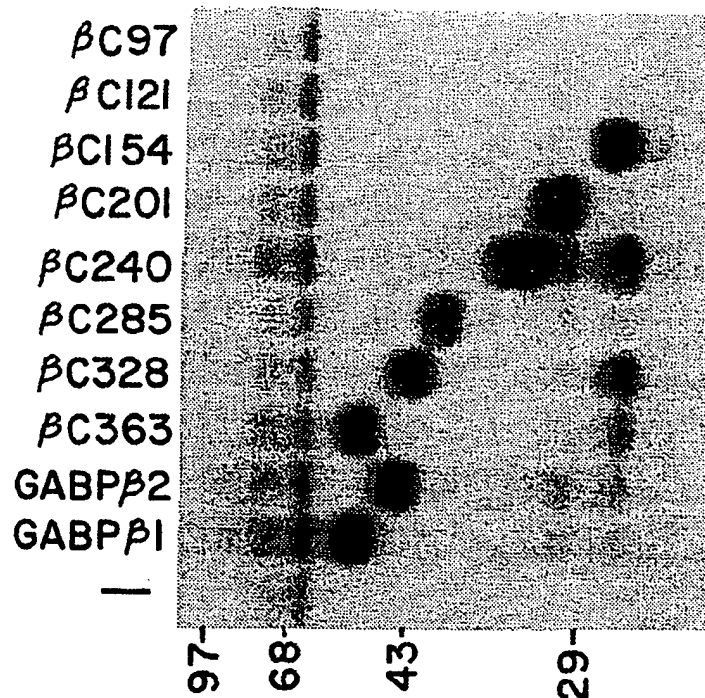


FIG. 12A2

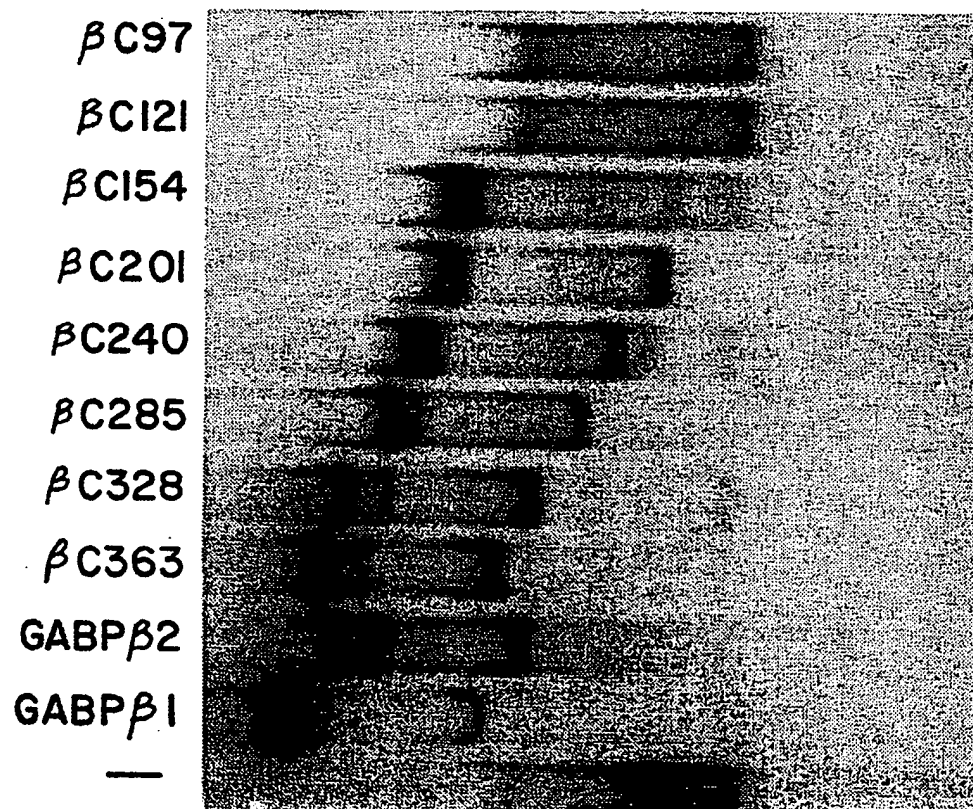


FIG. 12 B1

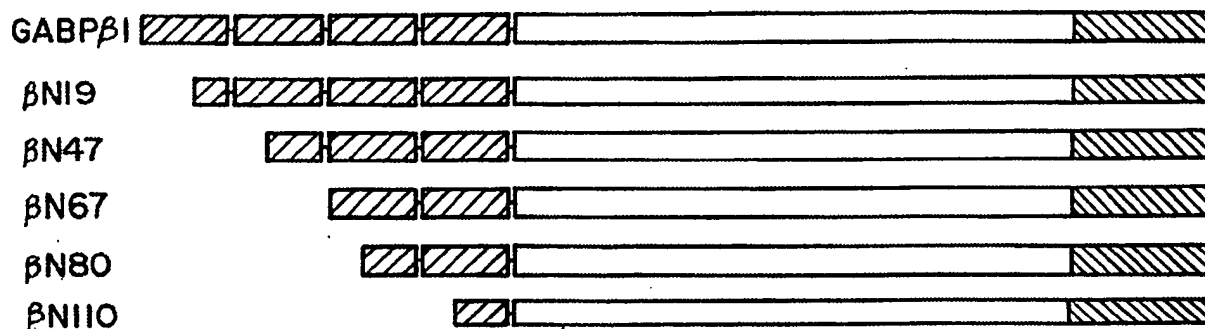


FIG. 12 B2

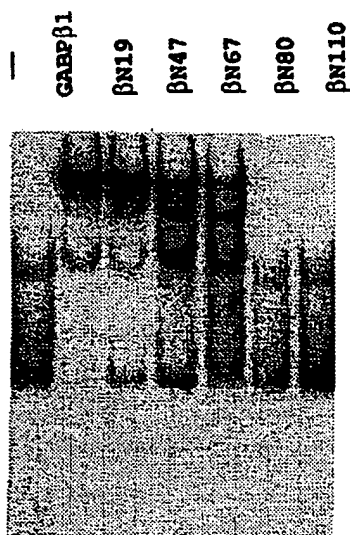


FIG. 12 B3

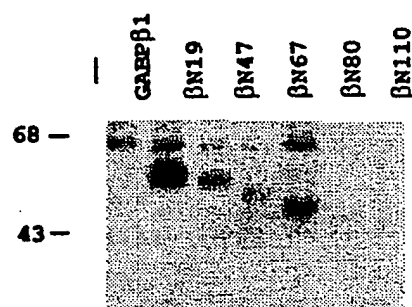
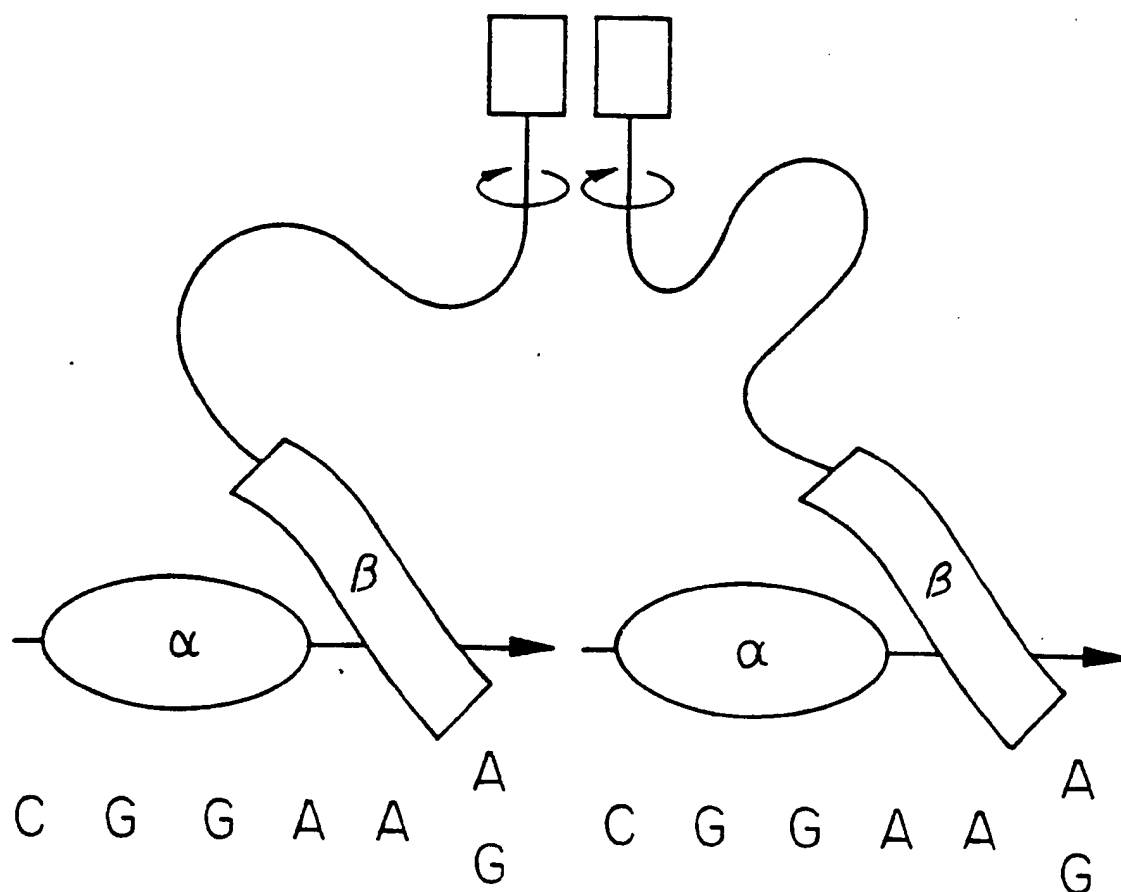


FIG. 13





# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US92/06748

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) : C12N 15/00, 15/12, 15/63, 15/70, 15/74, 15/79

US CL : 435/69.1, 172.3, 252.3, 320.1; 536/27; 935/10, 11, 27, 70, 72

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.1, 172.3, 252.3, 320.1; 536/27; 935/10, 11, 27, 70, 72

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
Please See Extra Sheet.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	GENES AND DEVELOPMENT, Volume 3, issued 1989, K.L. LaMarco et al, "Purification of a set of cellular polypeptides that bind to the purine-rich <u>cis</u> -regulatory element of herpes simplex virus immediate early genes", pages 1372-1383, see entire document.	1-13
Y	CELL, Volume 62, issued 07 September 1990, S. Ghosh et al, "Cloning of the p50 binding subunit of NF-kB: Homology to <u>rel</u> and <u>dorsal</u> ", pages 1019-1029, see entire document.	1-13
Y	THE EMBO JOURNAL, Volume 7, No. 3, issued 1988, K.E. Boulukos et al, "Identification of an evolutionarily conserved cellular <u>ets-2</u> gene ( <u>c-ets-2</u> ) encoding nuclear proteins related to the products of the <u>c-ets</u> proto-oncogene", pages 697-705, see entire document.	1-3, 6 and 9-13

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	* T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
* A document defining the general state of the art which is not considered to be part of particular relevance	* X	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
* E earlier document published on or after the international filing date	* Y	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
* L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	* Z	document member of the same patent family
* O document referring to an oral disclosure, use, exhibition or other means		
* P document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

14 NOVEMBER 1992

Date of mailing of the international search report

35 NOV 1992

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US92/06748

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	THE JOURNAL OF BIOLOGICAL CHEMISTRY, Volume 265, No. 14, issued 15 May 1990, J.K. Nyborg et al, "Interaction of cellular proteins with the human T-cell leukemia virus type I transcriptional control region", pages 8230-8236, see entire document.	1-3 and 6
Y	SCIENCE, Volume 249, issued September 1990, C. Coffman et al, " <u>Xotch</u> , the <u>Xenopus</u> homolog of <u>Drosophila</u> <u>Notch</u> ", pages 1438-1441, see entire document.	4, 5 and 7-12
Y	ONCOGENE RESEARCH, Volume 3, issued 1988, E.S.P. Reddy et al, "Structure, expression and alternative splicing of the human c- <u>ets</u> -1 proto-oncogene", pages 239-246, see pages 240-245.	4, 5 and 7-13
Y	WO, A, 90/05745 (Habener et al) 31 May 1990, pages 14-27.	9-13
Y	CELL, Volume 61, No. 1, issued 6 April 1990, M.J. Klemsz et al, "The macrophage and B cell-specific transcription factor PU.1 is related to the <u>ets</u> oncogene", pages 113-124, see entire document.	1-11
Y	ONCOGENE, Volume 5, issued 1990, D. Ray et al, "The human homologue of the putative proto-oncogene Spi-1: Characterization and expression in tumors", pages 663-668, see entire document.	1-11
Y	JOURNAL OF VIROLOGY, Volume 65, No. 1, issued January 1991, R. Paul et al, "The <u>Sfpi</u> -1 proviral integration site of Friend erythroleukemia encodes the <u>ets</u> -related transcription factor Pu.1", pages 464-467, see entire document.	1-11
A	WO, A, 91/07423 (Liao et al) 30 May 1991, pages 12-39.	8-13
A	WO, A, 89/09777 (Sukhatme) 19 October 1989, pages 9-22.	9-11 and 13
A	GENES AND DEVELOPMENT, Volume 2, issued 1988, S.J. Triezenberg et al, "Evidence of DNA:protein interactions that mediate HSV-1 immediate early gene activation by VP16", pages 730-742, see pages 735-741.	1, 9-11
A	CELL, Volume 60, issued 23 March 1990, A.M. Spence et al, "The product of <u>fem</u> -1, a nematode sex-determining gene, contains a motif found in cell cycle control proteins and receptors for cell-cell interactions", pages 981-990, see pages 982-988.	1, 9-12
A	NATURE, Volume 348, issued 01 November 1990, V. Bours et al, "Cloning of a mitogen-inducible gene encoding a kB DNA-binding protein with homology to the <u>rel</u> oncogene and to cell-cycle motifs", pages 76-80, see entire document.	1, 9-11
A	MOLECULAR AND CELLULAR BIOLOGY, Volume 8, No. 5, O.M. Andrisani et al, "Three sequence-specific DNA-protein complexes are formed with the same promoter element essential for expression of the rat somatostatin gene", pages 1947-1956, see pages 1948-1955.	1, 9-12
A	SCIENCE, Volume 245, issued 28 July 1989, P.J. Mitchell et al, "Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins", pages 371-378, see entire document.	1
A	THE JOURNAL OF BIOLOGICAL CHEMISTRY, Volume 265, No. 36, issued 25 December 1990, J. Ozer et al, "Isolation and characterization of a cDNA clone for the CCAAT transcription factor EFl <sub>A</sub> reveals a novel structural motif", pages 22143-22152, see pages 22144-22152.	1, 9-11
A	GENES AND DEVELOPMENT, Volume 4, issued 1990, C.-M. Fan et al, "A DNA-	1, 9-11

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US92/06748

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, USA, Volume 88, issued February 1991, R. Meyer et al. "Cloning of the DNA-binding subunit of human nuclear factor kB: The level of its mRNA is strongly regulated by phorbol ester or tumor necrosis factor $\alpha$ ", pages 966-970, see entire document.	1-2 and 9-11
A,P	EP, A, 0,449,170 (Hoppe-Seyler et al) issued 02 October 1991, pages 4-7.	1, 9-11
A	GENES AND DEVELOPMENT, Volume 4, issued 1990, C.V. Gunther et al, "Sequence-specific DNA binding of the proto-oncoprotein <u>ets-1</u> defines a transcriptional activator sequence within the long terminal repeat of the Moloney murine sarcoma virus", pages 667-679.	1, 9-11
Y	JOURNAL OF VIROLOGY, Volume 62, No. 9, issued September 1988, D. LePrince et al, "Alternative splicing with the chicken c- <u>ets-1</u> locus: Implications for transduction within the E26 retrovirus of the c- <u>ets</u> proto-oncogene", pages 3233-3241, see especially pages 3235-3240 and Figure 7.	4-5, 7-8

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US92/06748

### B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

STN/Chemical Abstracts; DIALOG/Biosis. Derwent Biotechnology

Abstracts, Current Biotechnology Abstracts, Agricola, Medline;

EMBL-NEW 7; UEMBL 31-72; GenBank-NEW 7; GenBank 72; N-GeneSeq 7.

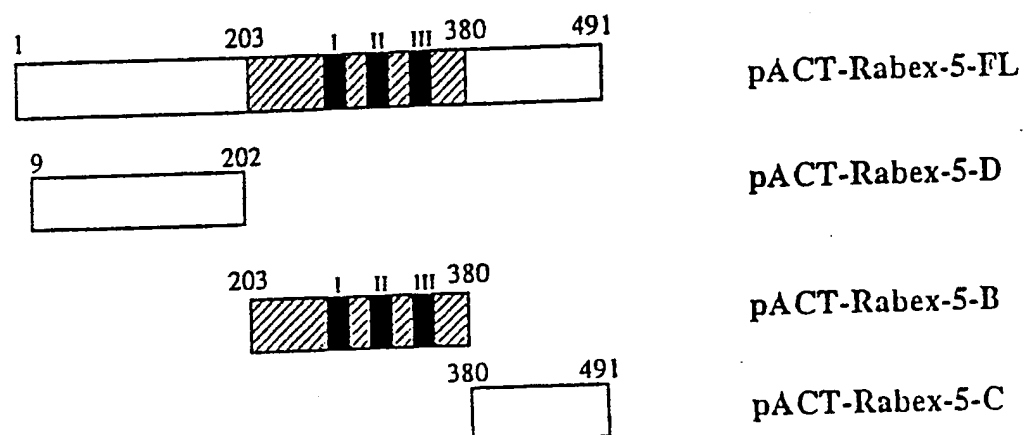
PIR 32; SwissProt 22; A-GeneSeq 7

FIG. 11A

FIG. 11B

10 20 30 40  
TNCCAANAANATATTCAAGTNTTTTTGGGGGGGCNAAATT 40  
CACCCNTNCGGTTAAAGGAACCAATTGATNTTTTTTTAA 80  
CAAAACCNCGGAGTTTTTCCTTNCGGGGGGGGTAAATANG 120  
GGGNAACCCAAAGATTNTTGCNATTCAATGCACAGNGGG 160  
ATGTNAACTAAAAACGGAGTTAAATATTTTAGGGGGGTTG 200  
210 220 230 240  
ACAGCAACCTGCATNGTAGAACCTTTTTTTTTNTTTCNGNG 240  
GACNTTNTATAACNTAAATATACCATTGATGATTTTNTTC 280  
CATTCAGTGACATCCACAGATTANGCAGCTATACTTGTGA 320  
AATCGTGCATGAGGCCCCCAGGGCACCGTTTTAGAACAAAC 360  
GTCACCTCACACAGGCAGGTGAGAAAGGTTCTCTTGCTTT 400  
410 420 430 440  
TCCAGTATCTTCCTAAGGATGGAGCCCAAATTCAGAGC 440  
AGTAACTTTGGAATAAAACCAGGGTGGGTATAAACTTCT 480  
TATTCTTAAATTTACATATAAGATCTATTAAGCTTGACAC 520  
ATCTGTGTCATCACGCACTGAAGACAGGAAGCAGTTCACT 560  
GAGTCAGCTGGTTCCCAAGCTCGCACAGAAGGTGATAAGT 600  
610 620 630 640  
TACTATCAAATGCCAGTGAGAATCTTCTTATAGAATAACC 640  
TGGGCCCCAAGTGATTTTAGTACAAAACCTTGCCCTTCTTG 680  
GTTTAATTTTCTATGTGCTTTTAGGTGTGAATCCAGATAT 720  
GCGGTCTTAATTCCTTTGGAAATACACAGTTCGTTTAGTT 760  
ACTGTACACTCTGTTTGTTCATAAACTGCATATCAACTT 800  
810 820 830 840  
CCAAAAAAAAAAAAAAAAAAAAA 822

FIG. 11C



DNA-binding hybrid	Activation domain hybrid	Colony color	Relative β-Gal units
pAS2-mHRas	pACT-Rabex-5-FL	Blue	229
pAS2-mHRas	pACT-Rabex-5-D	Blue	14
pAS2-mHRas	pACT-Rabex-5-B	Blue	24
pAS2-mHRas	pACT-Rabex-5-C	Blue	138
pAS2-yRas2p	pACT-Rabex-5-FL	White	6
pAS2-yRas2p	pACT-Rabex-5-D	White	4
pAS2-yRas2p	pACT-Rabex-5-B	White	<1
pAS2-yRas2p	pACT-Rabex-5-C	White	<1

FIG. 12



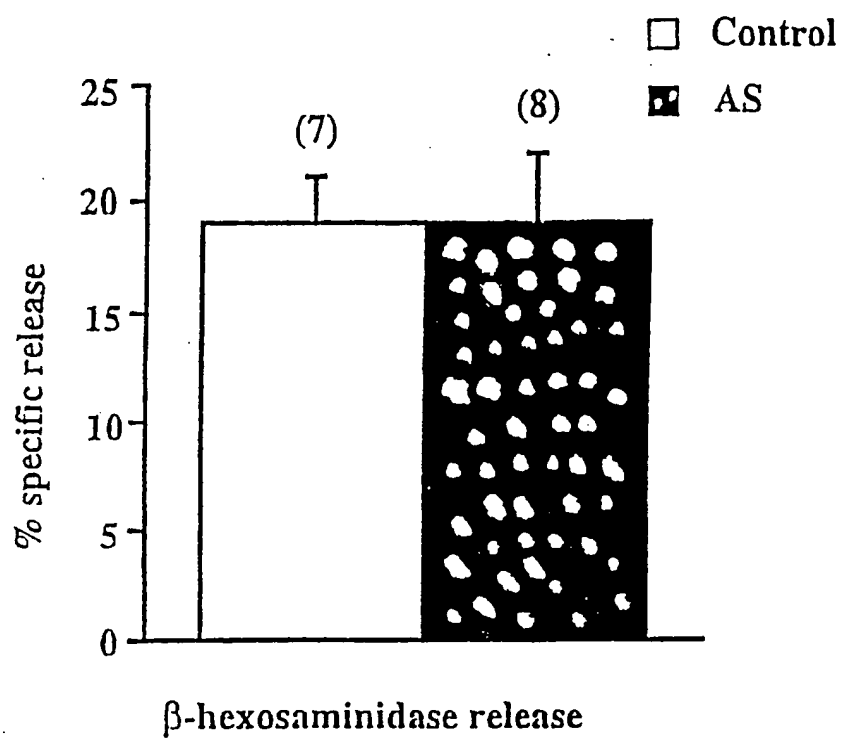


FIG. 13A

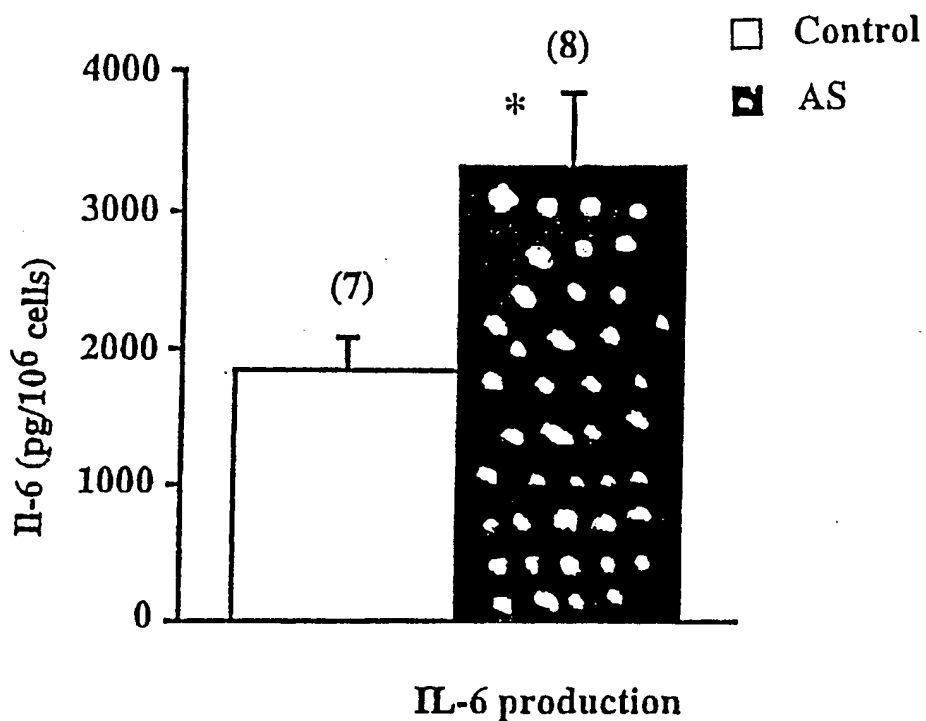


FIG. 13B

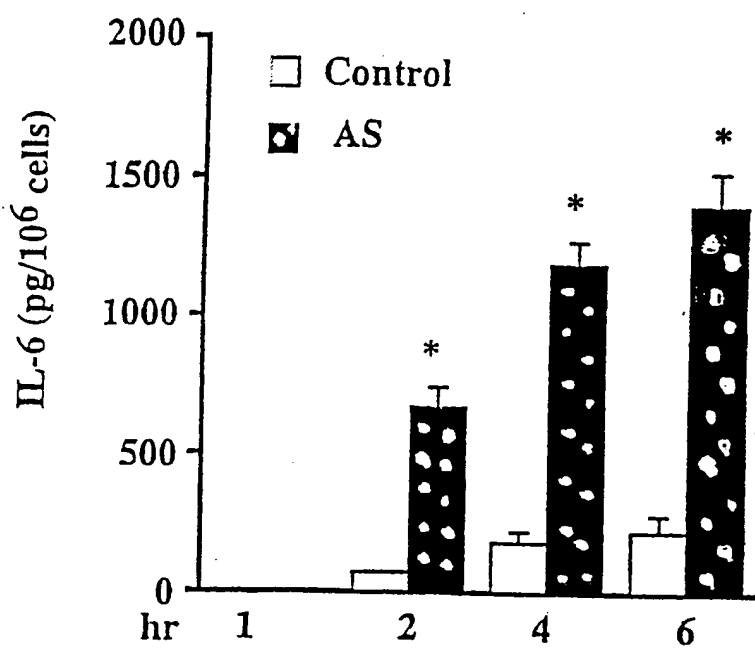


FIG. 13C

18 / 18

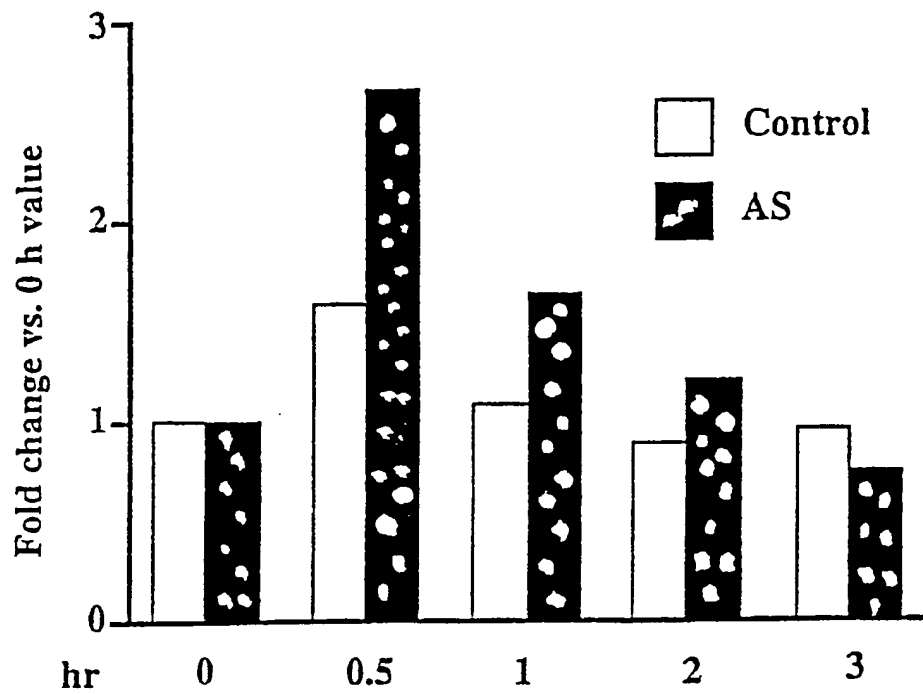


FIG. 14

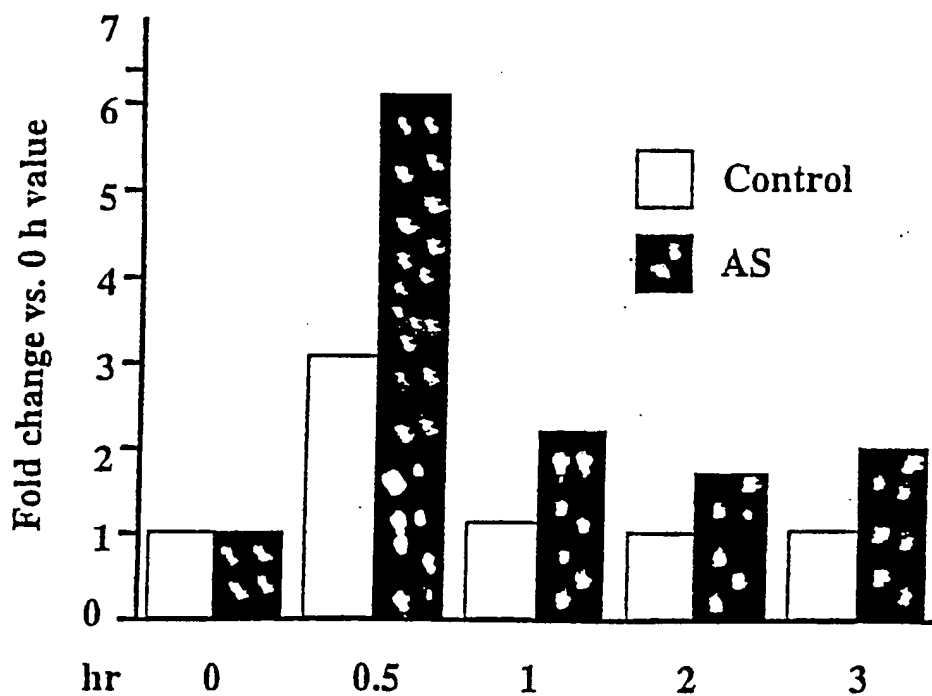


FIG. 15

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 98/19056

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/12 C07K14/47 C07K16  
G01N33/53

C12Q1/68 G01N33/50

According to International Patent Classification (IPC) or to both national class

and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
IPC 6 C12N C07K C12Q G01N

Documentation searched other than minimum documentation to the extent the

documents are included in the fields searched

Electronic data base consulted during the international search (name of data

base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 666 314 A (COLD SPRING HARBOR LAB) 9 August 1995 SEQ ID NOS: 13-16 see page 12, line 53 - line 56	9, 10, 12, 15, 16
X	J. COLICELLI ET AL.: "Expression of three mammalian cDNAs that interfere with RAS function in Saccharomyces cerevisiae" PROC. NATL. ACAD. SCI., vol. 88, April 1991, pages 2913-2917, XP002087363 NATL. ACAD. SCI., WASHINGTON, DC, US; see the whole document	9, 10, 12, 15, 16

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

## \* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"Z" document member of the same patent family

Date of the actual completion of the international search

Date of mailing of the international search report

10 December 1998

28/12/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax (+31-70) 340-2040

Authorized officer:

HORNIG H

## INTERNATIONAL SEARCH REPORT

Int'l. Application No

PCT/US 98/19056

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>L. HAN ET AL.: "Protein binding and signalling properties of RIN1 suggest a unique effector function"</p> <p>PROC. NATL. ACAD. SCI., vol. 94, May 1997, pages 4954-4959, XP002087364</p> <p>NATL. ACAD. SCI., WASHINGTON, DC, US; cited in the application see the whole document</p> <p>---</p>	9,10,12, 15,16
X	<p>C.G. BURD ET AL.: "A yeast protein related to a mammalian ras-binding protein, Vps9p, is required for localization of vacuolar proteins"</p> <p>MOL. CELL. BIOL., vol. 16, no. 5, May 1996, pages 2369-2377, XP002087365</p> <p>ASM WASHINGTON, DC, US cited in the application see the whole document</p> <p>---</p>	9,10,12, 15,16
A	<p>H. STENMARK ET AL.: "Rabaptin-5 is a direct effector of the small GTPase Rab5 in endocytic membrane fusion"</p> <p>CELL, vol. 83, 3 November 1995, pages 423-432, XP002087366</p> <p>CELL PRESS, CAMBRIDGE, MA, US; see the whole document</p> <p>---</p>	1-23
P, X	<p>HORIUCHI H ET AL: "A novel Rab5 GDP/GTP exchange factor complexed to Rabaptin-5 links nucleotide exchange to effector recruitment and function."</p> <p>CELL, (1997 SEP 19) 90 (6) 1149-59. JOURNAL CODE: CQ4. ISSN: 0092-8674., XP002087367</p> <p>see the whole document</p> <p>-----</p>	1-8,11, 13,14, 17-21

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 98/19056

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☒ Claims Nos.: 24-54  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:  
Please see Further Information sheet PCT/ISA/210
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

# INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No

PCT/US 98/19056

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0666314 A	09-08-1995	CA 2080920 A	21-10-1991
		EP 0537173 A	21-04-1992
		WO 9116457 A	31-10-1991
		US 5527896 A	18-06-1996

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Claims Nos.: 24-54

Claims 24 relating to agents which alter Rin2 activity relating to claims 25-57 could not be searched as the subject matters were not sufficiently disclosed.

Furthermore Claims 9,10,12,15,16, make no sense SEQ ID NOS.  
12,13,14,15 are not nucleotide sequences but amino acids